```
Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    213 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 254


RESULT 10
US-07-772-979-1
; Sequence 1, Application US/07772979
;  GENERAL INFORMATION:
;    APPLICANT:  Mizoguchi, Junzo
;    APPLICANT:  Nogawa, Makoto
;    APPLICANT:  Yamashita, Tomoyuki
;    APPLICANT:  Kubota, Akiko
;    APPLICANT:  Maeda, Tadashi
;    APPLICANT:  Taniuchi, Masahiko
;    APPLICANT:  Kobayashi, Junji
;    APPLICANT:  Satoh, Sakae
;    APPLICANT:  Otani, Masaru
;    TITLE OF INVENTION:  STABLE PRODUCTION OF M-CSF ACTIVE
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/772,979
;      FILING DATE:  19911008
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, Norman F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2103-025-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  693 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  M1-del
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..690
US-07-772-979-1


  Query Match             100.0%;  Score 42;  DB 3;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 0.0001;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 92
```

```
US-08-426-571-3
; Sequence 3, Application US/08426571
;  GENERAL INFORMATION:
;    APPLICANT:  LADNER, MARTHA B.
;    APPLICANT:  NOBLE, JANELLE A.
;    APPLICANT:  MARTIN, GEORGE A.
;    APPLICANT:  KAWASAKI, ERNEST S.
;    APPLICANT:  COYNE, MAZIE YEE
;    APPLICANT:  HALENBECK, ROBERT F.
;    APPLICANT:  KOTHS, KIRSTON E.
;    TITLE OF INVENTION:  NEW FORMS OF COLONY STIMULATING FACTOR-1
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  Intellectual Property - R440, P.O. Box 8097
;      CITY:  Emveryville
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,571
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGarrigle Jr., Philip L.
;      REGISTRATION NUMBER:  31,395
;      REFERENCE/DOCKET NUMBER:  0681.015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2718
;      TELEFAX:  (510) 655-3542
;      TELEX:  n/a
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1642 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  179..946
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  275..946
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  340
;      OTHER INFORMATION:  /note= "Intron Sequence"
US-08-426-571-3
```

→ USPN 6,156,300

```
  Query Match             100.0%;  Score 42;  DB 8;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT 12
US-60-278-258-1861
; Sequence 1861, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
```

```
Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    213 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 254


RESULT  10
US-07-772-979-1
; Sequence 1, Application US/07772979
;  GENERAL INFORMATION:
;    APPLICANT:  Mizoguchi, Junzo
;    APPLICANT:  Nogawa, Makoto
;    APPLICANT:  Yamashita, Tomoyuki
;    APPLICANT:  Kubota, Akiko
;    APPLICANT:  Maeda, Tadashi
;    APPLICANT:  Taniuchi, Masahiko
;    APPLICANT:  Kobayashi, Junji
;    APPLICANT:  Satoh, Sakae
;    APPLICANT:  Otani, Masaru
;    TITLE OF INVENTION:  STABLE PRODUCTION OF M-CSF ACTIVE
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/772,979
;      FILING DATE:  19911008
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, Norman F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2103-025-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  693 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  M1-del
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..690
US-07-772-979-1


  Query Match             100.0%;  Score 42;  DB 3;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 0.0001;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 CTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 92
```

```
US-08-426-571-3
; Sequence 3, Application US/08426571
;  GENERAL INFORMATION:
;    APPLICANT:  LADNER, MARTHA B.
;    APPLICANT:  NOBLE, JANELLE A.
;    APPLICANT:  MARTIN, GEORGE A.
;    APPLICANT:  KAWASAKI, ERNEST S.
;    APPLICANT:  COYNE, MAZIE YEE
;    APPLICANT:  HALENBECK, ROBERT F.
;    APPLICANT:  KOTHS, KIRSTON E.
;    TITLE OF INVENTION:  NEW FORMS OF COLONY STIMULATING FACTOR-1
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  Intellectual Property - R440, P.O. Box 8097
;      CITY:  Emveryville
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,571
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGarrigle Jr., Philip L.
;      REGISTRATION NUMBER:  31,395
;      REFERENCE/DOCKET NUMBER:  0681.015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2718
;      TELEFAX:  (510) 655-3542
;      TELEX:  n/a
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1642 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  179..946
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  275..946
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  340
;      OTHER INFORMATION:  /note= "Intron Sequence"
US-08-426-571-3


  Query Match             100.0%;  Score 42;  DB 8;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  12
US-60-278-258-1861
; Sequence 1861, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
```

USPN 6,156,300

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:52:43 ; Search time 2666.35 Seconds
(without alignments)
340.790 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgttgttggtct.....ctcctggcgagcaggagtat 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	100.0	42	1	PCT-US99-20344-11	Sequence 11, Appl
2	42	100.0	42	30	US-09-786-214-11	Sequence 11, Appl
3	42	100.0	60	1	PCT-US99-20344-8	Sequence 8, Appli
4	42	100.0	60	30	US-09-786-214-8	Sequence 8, Appli
5	42	100.0	78	1	PCT-US99-20344-4	Sequence 4, Appli
6	42	100.0	78	30	US-09-786-214-4	Sequence 4, Appli
7	42	100.0	129	1	PCT-US99-20344-50	Sequence 50, Appl
8	42	100.0	129	30	US-09-786-214-50	Sequence 50, Appl
9	42	100.0	609	29	US-09-758-449-148	Sequence 148, App
10	42	100.0	693	3	US-07-772-979-1	Sequence 1, Appli
11	42	100.0	1642	8	US-08-426-571-3	Sequence 3, Appli
12	42	100.0	1690	66	US-60-278-258-1861	Sequence 1861, Ap
13	42	100.0	1994	37	US-10-044-090-566	Sequence 566, App
14	42	100.0	1994	65	US-60-260-483-566	Sequence 566, App
15	42	100.0	1994	66	US-60-278-258-1859	Sequence 1859, Ap
16	42	100.0	2237	3	US-07-939-295-1	Sequence 1, Appli
17	42	100.0	2482	56	US-60-172-373-15186	Sequence 15186, A
18	42	100.0	2771	66	US-60-278-258-1863	Sequence 1863, Ap
19	42	100.0	3391	58	US-60-195-106-52	Sequence 52, Appl
20	42	100.0	3392	59	US-60-209-106-415	Sequence 415, App
21	42	100.0	3640	56	US-60-172-373-1570	Sequence 1570, Ap
22	42	100.0	3939	18	US-09-442-589B-619	Sequence 619, App
23	42	100.0	6306	4	US-08-004-141A-1	Sequence 1, Appli
24	40	95.2	2302	8	US-08-426-571-1	Sequence 1, Appli
25	32.4	77.1	320	28	US-09-704-424-28321	Sequence 28321, A
26	32.4	77.1	338	28	US-09-704-424-29722	Sequence 29722, A
27	32.4	77.1	369	28	US-09-704-424-24902	Sequence 24902, A
28	32.4	77.1	1987	8	US-08-426-571-23	Sequence 23, Appl
29	32.4	77.1	3920	75	US-60-360-207-11684	Sequence 11684, A
30	32.4	77.1	3931	8	US-08-426-571-21	Sequence 21, Appl
31	30.8	73.3	319	28	US-09-704-424-30476	Sequence 30476, A

	32	25	59.5	364	31	US-09-821-837-5558	Sequence 5558, Ap
	33	25	59.5	479	25	US-09-649-163-7958	Sequence 7958, Ap
	34	25	59.5	479	25	US-09-652-911-7860	Sequence 7860, Ap
	35	25	59.5	488	18	US-09-474-436-629	Sequence 629, App
	36	25	59.5	488	25	US-09-649-163-5318	Sequence 5318, Ap
	37	25	59.5	489	18	US-09-474-436-237	Sequence 237, App
	38	25	59.5	489	25	US-09-649-163-5283	Sequence 5283, Ap
	39	25	59.5	489	25	US-09-652-911-1444	Sequence 1444, Ap
	40	25	59.5	495	18	US-09-474-436-6152	Sequence 6152, Ap
	41	25	59.5	644	25	US-09-649-163-10225	Sequence 10225, A
	42	25	59.5	206025	63	US-60-243-468-603	Sequence 603, App
	43	24.6	58.6	451	33	US-09-873-402A-36830	Sequence 36830, A
	44	24.6	58.6	451	59	US-60-209-830-36692	Sequence 36692, A
c	45	23.4	55.7	5061	24	US-09-620-392-23282	Sequence 23282, A

ALIGNMENTS

```
RESULT   1
PCT-US99-20344-11
; Sequence 11, Application PC/TUS9920344
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
;  TITLE OF INVENTION: HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: PCT/US99/20344
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 11
;   LENGTH: 42
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..42
PCT-US99-20344-11


  Query Match           100.0%;  Score 42;  DB 1;  Length 42;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42


RESULT   2
US-09-786-214-11
; Sequence 11, Application US/09786214
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
;  TITLE OF INVENTION:  HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: US/09/786,214
;  CURRENT FILING DATE:  2001-06-14
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 11
;   LENGTH: 42
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..42
US-09-786-214-11


  Query Match           100.0%;  Score 42;  DB 30;  Length 42;
  Best Local Similarity  100.0%;  Pred. No. 7.5e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42


RESULT   3
PCT-US99-20344-8
; Sequence 8, Application PC/TUS9920344
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
;  TITLE OF INVENTION: HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: PCT/US99/20344
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 8
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..60
PCT-US99-20344-8


  Query Match           100.0%;  Score 42;  DB 1;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51


RESULT   4
US-09-786-214-8
; Sequence 8, Application US/09786214
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
;  TITLE OF INVENTION:  HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: US/09/786,214
;  CURRENT FILING DATE:  2001-06-14
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 8
;   LENGTH: 60
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..60
US-09-786-214-8


  Query Match           100.0%;  Score 42;  DB 30;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 7.8e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51


RESULT   5
PCT-US99-20344-4
; Sequence 4, Application PC/TUS9920344
```

```
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
;  TITLE OF INVENTION: HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: PCT/US99/20344
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 4
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..75
PCT-US99-20344-4


  Query Match             100.0%;  Score 42;  DB 1;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 8e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51


RESULT   6
US-09-786-214-4
; Sequence 4, Application US/09786214
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
;  TITLE OF INVENTION:  HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: US/09/786,214
;  CURRENT FILING DATE:  2001-06-14
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 4
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..75
US-09-786-214-4


  Query Match             100.0%;  Score 42;  DB 30;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 8e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51


RESULT   7
PCT-US99-20344-50
; Sequence 50, Application PC/TUS9920344
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FRAME
;  TITLE OF INVENTION: HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: PCT/US99/20344
;  CURRENT FILING DATE: 1999-09-03
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 50
;   LENGTH: 129
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..129
PCT-US99-20344-50


  Query Match             100.0%;  Score 42;  DB 1;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT   8
US-09-786-214-50
; Sequence 50, Application US/09786214
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research
;  TITLE OF INVENTION: AN ANTIGENIC PEPTIDE ENCODED BY AN ALTERNATIVE OPEN READING FR
;  TITLE OF INVENTION:  HUMAN MACROPHAGE COLONY-STIMULATING FACTOR
;  FILE REFERENCE: L0461/7040WO
;  CURRENT APPLICATION NUMBER: US/09/786,214
;  CURRENT FILING DATE:  2001-06-14
;  NUMBER OF SEQ ID NOS: 51
;  SOFTWARE: FastSEQ for Window Version 3.0
; SEQ ID NO 50
;   LENGTH: 129
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..129
US-09-786-214-50


  Query Match             100.0%;  Score 42;  DB 30;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT   9
US-09-758-449-148
; Sequence 148, Application US/09758449
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PM026
;  CURRENT APPLICATION NUMBER: US/09/758,449
;  CURRENT FILING DATE:  2001-01-11
;  PRIOR APPLICATION NUMBER: 60/179,065
;  PRIOR FILING DATE: 2000-01-31
;  PRIOR APPLICATION NUMBER: 60/180,628
;  PRIOR FILING DATE: 2000-02-04
;  NUMBER OF SEQ ID NOS: 1478
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 148
;   LENGTH: 609
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-758-449-148


  Query Match             100.0%;  Score 42;  DB 29;  Length 609;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-05;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    213 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 254


RESULT  10
US-07-772-979-1
; Sequence 1, Application US/07772979
;  GENERAL INFORMATION:
;    APPLICANT:  Mizoguchi, Junzo
;    APPLICANT:  Nogawa, Makoto
;    APPLICANT:  Yamashita, Tomoyuki
;    APPLICANT:  Kubota, Akiko
;    APPLICANT:  Maeda, Tadashi
;    APPLICANT:  Taniuchi, Masahiko
;    APPLICANT:  Kobayashi, Junji
;    APPLICANT:  Satoh, Sakae
;    APPLICANT:  Otani, Masaru
;    TITLE OF INVENTION:  STABLE PRODUCTION OF M-CSF ACTIVE
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/772,979
;      FILING DATE:  19911008
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, Norman F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2103-025-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  693 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  M1-del
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..690
US-07-772-979-1


  Query Match             100.0%;  Score 42;  DB 3;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 0.0001;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 92


RESULT  11
```

```
US-08-426-571-3
; Sequence 3, Application US/08426571
;  GENERAL INFORMATION:
;    APPLICANT:  LADNER, MARTHA B.
;    APPLICANT:  NOBLE, JANELLE A.
;    APPLICANT:  MARTIN, GEORGE A.
;    APPLICANT:  KAWASAKI, ERNEST S.
;    APPLICANT:  COYNE, MAZIE YEE
;    APPLICANT:  HALENBECK, ROBERT F.
;    APPLICANT:  KOTHS, KIRSTON E.
;    TITLE OF INVENTION:  NEW FORMS OF COLONY STIMULATING FACTOR-1
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  Intellectual Property - R440, P.O. Box 8097
;      CITY:  Emveryville
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,571
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGarrigle Jr., Philip L.
;      REGISTRATION NUMBER:  31,395
;      REFERENCE/DOCKET NUMBER:  0681.015
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2718
;      TELEFAX:  (510) 655-3542
;      TELEX:  n/a
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1642 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  179..946
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  275..946
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  340
;      OTHER INFORMATION:  /note= "Intron Sequence"
US-08-426-571-3


  Query Match             100.0%;  Score 42;  DB 8;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  12
US-60-278-258-1861
; Sequence 1861, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
```

→ USPN 6,156,300

```
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms Identified Thereby
;  FILE REFERENCE: GX-0010-1 P
;  CURRENT APPLICATION NUMBER: US/60/278,258
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 17730
;  SOFTWARE: PERL Program
; SEQ ID NO 1861
;   LENGTH: 1690
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 333948.12
US-60-278-258-1861


  Query Match            100.0%;  Score 42;  DB 66;  Length 1690;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    470 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 511


RESULT  13
US-10-044-090-566
; Sequence 566, Application US/10044090
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 566
;   LENGTH: 1994
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 1985219CB1
US-10-044-090-566


  Query Match            100.0%;  Score 42;  DB 37;  Length 1994;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    207 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 248


RESULT  14
US-60-260-483-566
; Sequence 566, Application US/60260483
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 P
;  CURRENT APPLICATION NUMBER: US/60/260,483
;  CURRENT FILING DATE:  2000-12-21
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 566
;   LENGTH: 1994
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 1985219CB1
US-60-260-483-566


  Query Match            100.0%;  Score 42;  DB 65;  Length 1994;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    207 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 248


RESULT  15
US-60-278-258-1859
; Sequence 1859, Application US/60278258
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;  TITLE OF INVENTION:  Polynucleotide Sequence Databases, and Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms Identified Thereby
;  FILE REFERENCE: GX-0010-1 P
;  CURRENT APPLICATION NUMBER: US/60/278,258
;  CURRENT FILING DATE:  2001-03-23
;  NUMBER OF SEQ ID NOS: 17730
;  SOFTWARE: PERL Program
; SEQ ID NO 1859
;   LENGTH: 1994
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 333948.11
US-60-278-258-1859


  Query Match            100.0%;  Score 42;  DB 66;  Length 1994;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    207 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 248


Search completed: September 11, 2002, 05:37:54
Job time: 6311 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:46:47 ; Search time 57.95 Seconds
(without alignments)
23.214 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	61.1	381	2	F75270	cytochrome P450 -
2	43	59.7	211	2	A64490	precorrin-6Y methy
3	43	59.7	623	2	T40991	probable lysophosp
4	43	59.7	821	2	C84304	DNA helicase [impo
5	41	56.9	156	2	F87551	conserved hypothet
6	41	56.9	326	2	T45226	probable N5,N10-me
7	41	56.9	508	2	E70764	probable cobI prot
8	40	55.6	227	2	B90400	hypothetical prote
9	40	55.6	243	2	I54459	MHC H-2K1-k - mous
10	40	55.6	428	2	AG1304	uracil permease ho
11	40	55.6	428	2	AG1676	uracil permease ho
12	40	55.6	540	2	A75250	carboxylesterase,
13	40	55.6	661	2	G84511	hypothetical prote
14	40	55.6	673	2	T50281	probable lysophosp
15	39	54.2	43	2	S21065	Ig kappa chain V r
16	39	54.2	96	2	S45441	Ig kappa chain V r
17	39	54.2	103	2	S19975	Ig kappa chain V r
18	39	54.2	106	2	PS0070	Ig kappa chain V r
19	39	54.2	106	2	PC4282	Ig kappa chain (an
20	39	54.2	107	2	S57444	Ig kappa chain V-J
21	39	54.2	108	2	C30502	Ig kappa chain V r
22	39	54.2	108	2	S33988	Ig kappa chain V r
23	39	54.2	108	2	G44151	Ig kappa chain V r
24	39	54.2	111	2	S23628	Ig kappa chain V r
25	39	54.2	114	2	S54905	Ig kappa chain V r
26	39	54.2	115	1	K3HUVG	Ig kappa chain pre
27	39	54.2	115	1	KVMSL7	Ig kappa chain pre
28	39	54.2	115	2	S11697	Ig kappa chain pre
29	39	54.2	116	2	B25521	Ig kappa chain pre
30	39	54.2	119	2	S41816	Ig kappa chain V r
31	39	54.2	125	2	S40344	Ig kappa chain V-J
32	39	54.2	128	2	PN0445	Ig kappa chain pre
33	39	54.2	128	2	S40379	Ig kappa chain V-J
34	39	54.2	128	2	A56701	Ig kappa chain V r
35	39	54.2	129	2	S29627	Ig kappa chain V r
36	39	54.2	129	2	S40363	Ig kappa chain - h
37	39	54.2	132	2	S05268	Ig kappa chain pre
38	39	54.2	144	2	PL0106	Ig kappa chain pre
39	39	54.2	144	2	B30502	Ig heavy chain V r
40	39	54.2	215	2	A23746	Ig kappa chain V-I
41	39	54.2	277	2	F82130	formyltetrahydrofo
42	39	54.2	278	2	E64131	formyltetrahydrofo
43	39	54.2	319	2	AD0941	probable regulator
44	39	54.2	563	2	JQ0623	nerve growth facto
45	39	54.2	1367	2	S74285	BUD3 protein - yea

ALIGNMENTS

RESULT 1
F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g646
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 61.1%; Score 44; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSP 9
|||||||||
Db 52 LPAVVGLSP 60

RESULT 2
A64490
precorrin-6Y methylase homolog - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: A64490
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A;Reference number: A64300; MUID:96337999
A;Accession: A64490
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-211 <BUL>
A;Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99541.1; PID:g1592152
C;Genetics:
A;Map position: FOR1500322-1500957

C;Superfamily: precorrin-6Y methylase

Query Match 59.7%; Score 43; DB 2; Length 211;
Best Local Similarity 54.5%; Pred. No. 9.3;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 4 VVGLSPGEQEY 14
:||: ||::||
Db 4 IVGIGPGDREY 14

RESULT 3
T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T40991
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21962
A;Accession: T40991
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-623 <LYN>
A;Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GN00068; SPDB:SPCC1450.09c
A;Experimental source: strain 972h-; cosmid c1450
C;Genetics:
A;Gene: SPDB:SPCC1450.09c
A;Map position: 3
C;Superfamily: yeast lysophospholipase

Query Match 59.7%; Score 43; DB 2; Length 623;
Best Local Similarity 69.2%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
|| ||| |||| :
Db 76 PASDGLSTGEQEF 88

RESULT 4
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84304
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483
A;Accession: C84304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-821 <STO>
A;Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C;Genetics:
A;Gene: hel

Query Match 59.7%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGEQ 12
|||||||:||
Db 326 AVVGLSPAEQ 335

RESULT 5
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: F87551
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-156 <STO>
A;Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2439
C;Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match 56.9%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGEQ 12
||||| |||:
Db 18 AVVGLDPGEK 27

RESULT 6
T45226
probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [importe
C;Species: Methanolobus tindarius
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T45226
R;Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G
submitted to the EMBL Data Library, September 1998
A;Description: The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the f
A;Reference number: Z22947
A;Accession: T45226
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-326 <WES>
A;Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A;Experimental source: DSM 2278
C;Genetics:
A;Gene: ffdA

Query Match 56.9%; Score 41; DB 2; Length 326;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGEQ 12
|:::|| ||||
Db 88 AILGLGPGEQ 97

RESULT 7
E70764
probable cobI protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70764
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987

A;Accession: E70764
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-508 <COL>
A;Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: cobI

Query Match 56.9%; Score 41; DB 2; Length 508;
Best Local Similarity 58.3%; Pred. No. 51;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AVVGLSPGEQEY 14
||||| ||: ::
Db 250 AVVGLGPGDSDW 261

RESULT 8
B90400
hypothetical protein cbiE [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C;Accession: B90400
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <KUR>
A;Cross-references: GB:AE006641; NID:g13815599; PIDN:AAK42457.1; GSPDB:GN00155
C;Genetics:
A;Gene: cbiE
C;Superfamily: precorrin-6Y methylase

Query Match 55.6%; Score 40; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 33;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VVGLSPGEQEY 14
:||: ||: ||
Db 10 IVGVGPGDPEY 20

RESULT 9
I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

Query Match 55.6%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 35;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VGLSPGEQE 13
:|||||:|
Db 216 LGLSPGEEE 224

RESULT 10
AG1304
uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1304
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:g16411293; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 63;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
:|:| |||||
Db 42 VPSVTGLSPG 51

RESULT 11
AG1676
uracil permease homolog pyrP [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1676
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1676
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-428 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:g16414454; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: pyrP
C;Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 63;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
:|:| |||||
Db 42 VPSVTGLSPG 51

RESULT 12
A75250
carboxylesterase, type B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75250
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: A75250
A;Molecule type: DNA
A;Residues: 1-540 <WHI>
A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2626
A;Map position: 1
C;Superfamily: cholinesterase; cholinesterase homology

Query Match 55.6%; Score 40; DB 2; Length 540;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGE 11
| |:||:|||
Db 512 PQVLGLAPGE 521

RESULT 13
G84511
hypothetical protein At2g13900 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84511
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84511
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-661 <STO>
A;Cross-references: GB:AE002093; NID:g6598598; PIDN:AAF18650.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g13900
A;Map position: 2

Query Match 55.6%; Score 40; DB 2; Length 661;
Best Local Similarity 53.8%; Pred. No. 99;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
| : : ||||||
Db 437 PLTLKIVPGEQEY 449

RESULT 14
T50281
probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yea
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001
C;Accession: T50281; T42738
R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25053
A;Accession: T50281
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-673 <ZIM>
A;Cross-references: EMBL:AL137130; NID:g6742151; PIDN:CAB69631.1; PID:g6742159; GSPDB
A;Experimental source: strain 972h(-); cosmid c977
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722
A;Accession: T42738
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 427-673 <YOS>
A;Cross-references: EMBL:D89183; NID:g1749573; PIDN:BAA13845.1; PID:g1749574
A;Experimental source: strain PR745
C;Genetics:
A;Gene: SPDB:SPAC977.09c
A;Map position: 1
A;Introns: 651/3
C;Function:
A;Description: catalyzes the hydolysis of 2-lysophosphatidylcholine to glycerophospho
C;Superfamily: yeast lysophospholipase
C;Keywords: carboxylic ester hydrolase

Query Match 55.6%; Score 40; DB 2; Length 673;
Best Local Similarity 61.5%; Pred. No. 1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQEY 14
|| ||: ||| |
Db 83 PASEGLNEGEQSY 95

RESULT 15
S21065
Ig kappa chain V region (anti-RH(D)) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
C;Accession: S21065
R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.
Protein Seq. Data Anal. 4, 317-318, 1991
A;Title: Subgroup assignement of a human monoclonal anti-Rh(D) antibody.
A;Reference number: S21065; MUID:92253544
A;Accession: S21065
A;Molecule type: protein
A;Residues: 1-43 <DLO>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 54.2%; Score 39; DB 2; Length 43;
Best Local Similarity 63.6%; Pred. No. 8.5;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
|| : |||||:
Db 8 PATLSLSPGER 18

Search completed: September 9, 2002, 06:51:26
Job time: 279 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 02:45:43 ; Search time 1587.89 Seconds
(without alignments)
356.997 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgttgttggtct.....ctcctggcgagcaggagtat 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	42	100.0	250	10	BF848044	BF848044 CM3-EN004
	2	42	100.0	520	10	BF724253	BF724253 bx02f10.y
	3	42	100.0	540	9	AL556517	AL556517 AL556517
	4	42	100.0	974	10	BG424422	BG424422 602447739
	5	32.4	77.1	564	10	BE374641	BE374641 601224139
	6	32.4	77.1	579	10	BE912798	BE912798 601664754
	7	32.4	77.1	665	10	BI661774	BI661774 603304836
	8	32.4	77.1	672	10	BG974498	BG974498 602844291
	9	32.4	77.1	691	10	BI684603	BI684603 603306810
	10	32.4	77.1	731	10	BI657501	BI657501 603281837
	11	32.4	77.1	777	10	BI683670	BI683670 603306210
	12	32.4	77.1	825	10	BG975351	BG975351 602842993
	13	25.2	60.0	423	10	BG814544	BG814544 daf68f06.
	14	25.2	60.0	548	10	BJ028868	BJ028868 BJ028868
c	15	25.2	60.0	615	10	BJ049970	BJ049970 BJ049970
	16	24	57.1	579	10	BI789811	BI789811 ic43g08.y
c	17	23.8	56.7	653	10	BG779215	BG779215 602667168
	18	23.4	55.7	764	10	BE263821	BE263821 601194186
c	19	23.2	55.2	1135	10	BF783578	BF783578 602109577
	20	23	54.8	413	12	B40890	B40890 HS-1052-B1
c	21	23	54.8	661	10	BJ063201	BJ063201 BJ063201
	22	23	54.8	934	12	CNS03CRY	AL238183 Tetraodon
c	23	23	54.8	1569	12	AG070941	AG070941 Pan trog
c	24	22.8	54.3	282	12	AZ553140	AZ553140 RPCI-23-2
c	25	22.8	54.3	412	12	AZ702302	AZ702302 RPCI-23-2
	26	22.8	54.3	539	10	BF075725	BF075725 224983 MA
c	27	22.8	54.3	735	9	AI261290	AI261290 qz32c08.x
c	28	22.8	54.3	768	10	BI822715	BI822715 603036051
c	29	22.6	53.8	1137	10	BE786185	BE786185 601474304
c	30	22.4	53.3	230	12	AZ720631	AZ720631 RPCI-24-8
c	31	22.4	53.3	240	10	BI132004	BI132004 G129P30Y
c	32	22.4	53.3	327	10	BF497080	BF497080 AT11145.5
c	33	22.4	53.3	387	10	BE491273	BE491273 db70f08.y
c	34	22.4	53.3	406	10	BI127593	BI127593 G062P91Y
c	35	22.4	53.3	442	10	BI120788	BI120788 F022P83Y
c	36	22.4	53.3	611	9	AU244251	AU244251 AU244251
c	37	22.4	53.3	622	10	BF506570	BF506570 AT09090.5
c	38	22.4	53.3	627	10	BF490922	BF490922 AT27557.5
c	39	22.4	53.3	629	10	BF499014	BF499014 AT13494.5
c	40	22.4	53.3	637	10	BF495620	BF495620 AT04570.5
c	41	22.4	53.3	638	10	BG376989	BG376989 UI-R-CU0
c	42	22.4	53.3	683	10	BF488388	BF488388 AT23649.5
	43	22.4	53.3	704	12	AZ854829	AZ854829 2M0158D09
c	44	22.4	53.3	706	10	BF283113	BF283113 EST447704
c	45	22.4	53.3	753	10	BF505238	BF505238 AT07422.5

ALIGNMENTS

RESULT 1
BF848044
LOCUS BF848044 250 bp mRNA linear EST 16-JAN-2001
DEFINITION CM3-EN0048-031100-458-e08 EN0048 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF848044
VERSION BF848044.1 GI:12235194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM3&t2=CM3-EN0048-031100-458-e08&t3=2000-11-03&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 249.
FEATURES Location/Qualifiers
source 1. .250
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
                     /clone_lib="EN0048"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       47 a     51 c     89 g     61 t      2 others
ORIGIN


  Query Match             100.0%;  Score 42;  DB 10;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 0.00019;
  Matches   42;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    133 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 174


RESULT   2
BF724253
LOCUS       BF724253                520 bp    mRNA    linear   EST 05-JAN-2001
DEFINITION  bx02f10.y1 Human Iris cDNA (Un-normalized, unamplified): BX Homo
            sapiens cDNA clone bx02f10 5', mRNA sequence.
ACCESSION   BF724253
VERSION     BF724253.1  GI:12040162
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 02  row: f  column: 10
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="bx02f10"
                     /clone_lib="Human Iris cDNA (Un-normalized, unamplified):
                     BX"
                     /tissue_type="Iris"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris
                     tissue was pooled from 10 individuals ranging in age from
                     4-80 years and RNA was extracted. From this pooled sample
                     an aliquot of 60ug of total RNA yielded 2.17ug of mRNA. A
                     directionally cloned cDNA library in the pCMVSPORT6 vector
                     was constructed at Life Technologies, essentially
                     following the protocols of the SuperScript Plasmid System
                     full details of which are contained in the manufacturer's
                     Instruction manual (http://www.lifetech.com/). First
                     strand synthesis was carried out using a Not I
                     primer-adapter [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3'
                     ]. Not I/blunt end inserts were cloned into the Not I/EcoR
                     V sites in the vector. EST analysis was performed on the
                     unamplified library at the NIH Intramural Sequencing
                     Center (NISC)."
BASE COUNT      103 a    157 c    163 g     97 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 10;  Length 520;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    207 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 248


RESULT   3
AL556517
LOCUS       AL556517                540 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL556517 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK006YI21 5
            prime, mRNA sequence.
ACCESSION   AL556517
VERSION     AL556517.1  GI:12899263
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK006YI21"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      116 a    157 c    162 g    102 t      3 others
ORIGIN


  Query Match             100.0%;  Score 42;  DB 9;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    159 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 200


RESULT   4
BG424422
LOCUS       BG424422                974 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602447739F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4586312 5',
            mRNA sequence.
ACCESSION   BG424422
VERSION     BG424422.1  GI:13330928
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 974)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1314  row: n  column: 09
            High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1. .974
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4586312"
                     /clone_lib="NIH_MGC_14"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      199 a    279 c    299 g    196 t      1 others
ORIGIN


  Query Match           100.0%;  Score 42;  DB 10;  Length 974;
  Best Local Similarity  100.0%;  Pred. No. 0.00027;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||
Db    241 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 282


RESULT   5
BE374641
LOCUS       BE374641                564 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  601224139F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582536 5',
            mRNA sequence.
ACCESSION   BE374641
VERSION     BE374641.1  GI:9320004
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8737  row: n  column: 09
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3582536"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT;
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      152 a    153 c    138 g    121 t
ORIGIN


  Query Match           77.1%;  Score 32.4;  DB 10;  Length 564;
  Best Local Similarity  85.7%;  Pred. No. 0.43;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||||   ||||||||||||
Db     40 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 81


RESULT   6
BE912798
LOCUS       BE912798                579 bp    mRNA    linear   EST 29-SEP-2000
DEFINITION  601664754F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3964933 5',
            mRNA sequence.
ACCESSION   BE912798
VERSION     BE912798.1  GI:10409785
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 579)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9135  row: k  column: 14
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3964933"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT;
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      164 a    150 c    142 g    123 t
ORIGIN


  Query Match           77.1%;  Score 32.4;  DB 10;  Length 579;
```

```
  Best Local Similarity   85.7%;  Pred. No. 0.43;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    202 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 243


RESULT   7
BI661774
LOCUS       BI661774                665 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603304836F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350688 5',
            mRNA sequence.
ACCESSION   BI661774
VERSION     BI661774.1  GI:15576010
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11891  row: g  column: 09
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1. .665
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5350688"
                     /clone_lib="NIH_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2.5 kb.  Library constructed by Life
                     Technologies, catalog # 12018-016. Investigators providing
                     samples: Lothar Hennighausen/Priscilla Furth, NIH
                     Reference for transgenic model: Li et al., Cell Growth and
                     Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                     Library."
BASE COUNT      162 a    184 c    181 g    138 t
ORIGIN


  Query Match            77.1%;  Score 32.4;  DB 10;  Length 665;
  Best Local Similarity   85.7%;  Pred. No. 0.45;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    210 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 251


RESULT   8
BG974498
LOCUS       BG974498                672 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602844291F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979734 5',
            mRNA sequence.
ACCESSION   BG974498
VERSION     BG974498.1  GI:14362135
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 672)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10978  row: f  column: 23
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4979734"
                     /clone_lib="NIH_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2.5 kb.  Library constructed by Life
                     Technologies, catalog # 12018-016. Investigators providing
                     samples: Lothar Hennighausen/Priscilla Furth, NIH
                     Reference for transgenic model: Li et al., Cell Growth and
                     Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                     Library."
BASE COUNT      162 a    188 c    183 g    139 t
ORIGIN


  Query Match            77.1%;  Score 32.4;  DB 10;  Length 672;
  Best Local Similarity   85.7%;  Pred. No. 0.45;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    214 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 255


RESULT   9
BI684603
LOCUS       BI684603                691 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603306810F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5342906 5',
            mRNA sequence.
ACCESSION   BI684603
VERSION     BI684603.1  GI:15647231
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 691)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
            http://image.llnl.gov
            Plate: LLAM11871  row: c  column: 03
            High quality sequence stop: 686.
FEATURES              Location/Qualifiers
     source           1. .691
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5342906"
                      /clone_lib="NCI_CGAP_Mam6"
                      /sex="female, virgin"
                      /tissue_type="infiltrating ductal carcinoma"
                      /dev_stage="5 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      169 a    192 c    190 g    140 t
ORIGIN


  Query Match             77.1%;  Score 32.4;  DB 10;  Length 691;
  Best Local Similarity   85.7%;  Pred. No. 0.45;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    232 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 273


RESULT  10
BI657501
LOCUS       BI657501                731 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603281837F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5326185 5',
            mRNA sequence.
ACCESSION   BI657501
VERSION     BI657501.1  GI:15571737
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11827  row: j  column: 10
            High quality sequence stop: 731.
FEATURES              Location/Qualifiers
     source           1. .731
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5326185"
                      /clone_lib="NIH_CGAP_Mam4"
                      /tissue_type="tumor, gross tissue"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                      Average insert 2.5 kb.  Library constructed by Life
                      Technologies, catalog # 12018-016. Investigators providing
                      samples: Lothar Hennighausen/Priscilla Furth, NIH
                      Reference for transgenic model: Li et al., Cell Growth and
                      Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                      Library."
BASE COUNT      181 a    205 c    194 g    151 t
ORIGIN


  Query Match             77.1%;  Score 32.4;  DB 10;  Length 731;
  Best Local Similarity   85.7%;  Pred. No. 0.46;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    185 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 226


RESULT  11
BI683670
LOCUS       BI683670                777 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  603306210F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5342163 5',
            mRNA sequence.
ACCESSION   BI683670
VERSION     BI683670.1  GI:15646298
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1  (bases 1 to 777)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11869  row: d  column: 04
            High quality sequence stop: 765.
FEATURES              Location/Qualifiers
     source           1. .777
                      /organism="Mus musculus"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:5342163"
                      /clone_lib="NCI_CGAP_Mam6"
                      /sex="female, virgin"
                      /tissue_type="infiltrating ductal carcinoma"
                      /dev_stage="5 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigator
                      providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      208 a    211 c    203 g    155 t
ORIGIN


  Query Match             77.1%;  Score 32.4;  DB 10;  Length 777;
  Best Local Similarity   85.7%;  Pred. No. 0.47;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||| |||| || ||||||||||||   ||||||||||||
Db    348 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 389


RESULT  12
BG975351
LOCUS       BG975351                825 bp    mRNA    linear   EST 12-JUN-2001
```

```
DEFINITION  602842993F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4978679 5',
            mRNA sequence.
ACCESSION   BG975351
VERSION     BG975351.1  GI:14362988
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10975  row: j  column: 24
            High quality sequence stop: 689.
FEATURES             Location/Qualifiers
     source          1. .825
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4978679"
                     /clone_lib="NIH_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2.5 kb.  Library constructed by Life
                     Technologies, catalog # 12018-016. Investigators providing
                     samples: Lothar Hennighausen/Priscilla Furth, NIH
                     Reference for transgenic model: Li et al., Cell Growth and
                     Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                     Library."
BASE COUNT      190 a    233 c    231 g    171 t
ORIGIN
```

```
  Query Match             77.1%;  Score 32.4;  DB 10;  Length 825;
  Best Local Similarity   85.7%;  Pred. No. 0.47;
  Matches   36;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
           ||||| |||| || ||||||||||||||   ||||||||||||
Db     221 CTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTAT 262
```

```
RESULT  13
BG814544
LOCUS       BG814544                423 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  daf68f06.y1 NICHD XGC Eye1 Xenopus laevis cDNA clone IMAGE:4743706
            5' similar to TR:Q27924 Q27924 OVARY2. ;, mRNA sequence.
ACCESSION   BG814544
VERSION     BG814544.1  GI:14185524
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: daf68f06.x1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4743706"
                     /clone_lib="NICHD XGC Eye1"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: This is a Xenopus Gene Collection (XGC
                     ) library."
BASE COUNT      139 a     86 c    137 g     61 t
ORIGIN
```

```
  Query Match             60.0%;  Score 25.2;  DB 10;  Length 423;
  Best Local Similarity   78.9%;  Pred. No. 1.1e+02;
  Matches   30;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 cctgctgttgttggtctgtctcctggcgagcaggagta 41
           || || | || |||  |||||| ||||| |||||||||
Db     201 CCAGCGGGTGATGGAGTGTCTCATGGCGTGCAGGAGTA 238
```

```
RESULT  14
BJ028868
LOCUS       BJ028868                548 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  BJ028868 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL010j09 5', mRNA sequence.
ACCESSION   BJ028868
VERSION     BJ028868.1  GI:17398582
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL010j09"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
```

```
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      155 a    114 c    192 g     87 t
ORIGIN


  Query Match             60.0%;  Score 25.2;  DB 10;  Length 548;
  Best Local Similarity   78.9%;  Pred. No. 1.2e+02;
  Matches   30;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 cctgctgttgttggtctgtctcctggcgagcaggagta 41
          || || | || |||  |||||| ||||| |||||||||
Db    441 CCAGCGGGTGATGGAGTGTCTCATGGCGTGCAGGAGTA 478


RESULT  15
BJ049970/c
LOCUS       BJ049970                615 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ049970 NIBB Mochii normalized Xenopus neurula library Xenopus
            laevis cDNA clone XL021n02 3', mRNA sequence.
ACCESSION   BJ049970
VERSION     BJ049970.1  GI:17380018
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL021n02"
                     /clone_lib="NIBB Mochii normalized Xenopus neurula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 15"
BASE COUNT      123 a    183 c    128 g    179 t      2 others
ORIGIN


  Query Match             60.0%;  Score 25.2;  DB 10;  Length 615;
  Best Local Similarity   78.9%;  Pred. No. 1.2e+02;
  Matches   30;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 cctgctgttgttggtctgtctcctggcgagcaggagta 41
          || || | || |||  |||||| ||||| |||||||||
Db    615 CCAGCGGGTGATGGAGTGTCTCATGGCGTGCAGGAGTA 578


Search completed: September 11, 2002, 04:18:28
Job time: 5565 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:10:28 ; Search time 1829.97 Seconds
(without alignments)
480.289 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgttgttggtct.....ctcctggcgagcaggagtat 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	100.0	361	9	HUMCSF1M1	M11038 Human macro
2	42	100.0	693	6	E03655	E03655 DNA sequenc
3	42	100.0	821	6	I01437	I01437 Sequence 4
4	42	100.0	821	6	I09299	I09299 Sequence 3
5	42	100.0	931	9	HUMCSF1X	M76453 Homo sapien
6	42	100.0	1119	6	E12942	E12942 Human cDNA
7	42	100.0	1370	9	HSU22386	U22386 Human macro
8	42	100.0	1640	6	I08431	I08431 Sequence 4
9	42	100.0	1641	6	I00368	I00368 Sequence 1
10	42	100.0	1641	6	I01429	I01429 Sequence 1
11	42	100.0	1642	6	AR105651	AR105651 Sequence
12	42	100.0	1642	6	AR110609	AR110609 Sequence
13	42	100.0	1642	6	AR141545	AR141545 Sequence
14	42	100.0	1642	6	AR142870	AR142870 Sequence
15	42	100.0	1642	6	I01440	I01440 Sequence 7
16	42	100.0	1642	6	I08736	I08736 Sequence 2
17	42	100.0	1642	6	I09203	I09203 Sequence 1
18	42	100.0	1642	6	I09298	I09298 Sequence 1
19	42	100.0	1642	6	I27815	I27815 Sequence 3
20	42	100.0	1642	6	I28747	I28747 Sequence 3
21	42	100.0	1642	6	I50653	I50653 Sequence 3
22	42	100.0	1642	6	I67640	I67640 Sequence 3
23	42	100.0	1642	6	I71134	I71134 Sequence 3
24	42	100.0	1665	6	E12943	E12943 Human cDNA
25	42	100.0	1665	9	HUMMCSF	M64592 Human macro
26	42	100.0	1987	6	I02044	I02044 Sequence 1
27	42	100.0	1997	6	I03518	I03518 Sequence 1
28	42	100.0	2475	9	HUMMCSF1A	M27087 Homo sapien
29	42	100.0	2488	6	E01949	E01949 cDNA encodi
30	42	100.0	2488	6	E02182	E02182 DNA encodin
31	42	100.0	2488	23	E09614	E09614 Human cDNA
32	42	100.0	2500	6	E01950	E01950 cDNA encodi
33	42	100.0	2500	6	E02183	E02183 DNA encodin
34	42	100.0	2500	23	E09615	E09615 Human cDNA
35	42	100.0	2851	9	BC021117	BC021117 Homo sapi
36	42	100.0	3391	6	AX281643	AX281643 Sequence
37	42	100.0	3939	9	HUMCSDF1	M37435 Human macro
38	42	100.0	3981	6	I08586	I08586 Sequence 1
39	42	100.0	3981	6	I09309	I09309 Sequence 1
40	42	100.0	4021	6	E03219	E03219 DNA encodin
41	42	100.0	4021	6	E03235	E03235 cDNA encodi
42	41	97.6	814	6	I05804	I05804 Sequence 9
43	40	95.2	1664	6	I00370	I00370 Sequence 3
44	40	95.2	1844	6	I01431	I01431 Sequence 3
45	40	95.2	2302	6	AR105650	AR105650 Sequence

ALIGNMENTS

RESULT 1
HUMCSF1M1
LOCUS HUMCSF1M1 361 bp DNA linear PRI 08-AUG-1995
DEFINITION Human macrophage-specific colony-stimulating factor (CSF-1) gene, exon 1.
ACCESSION M11038
VERSION M11038.1 GI:181140
KEYWORDS alternative splicing; colony stimulating factor.
SEGMENT 1 of 3
SOURCE Human DNA and pancreatic tumor cell line MIA-PaCa-2, cDNA to mRNA, clone pcCSF-17.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 361)
AUTHORS Kawasaki,E.S., Ladner,M.B., Wang,A.M., Van Arsdell,J., Warren,M.K., Coyne,M.Y., Schweickart,V.L., Lee,M.-T., Wilson,K.J., Boosman,A., Stanley,E.R., Ralph,P. and Mark,D.F.
TITLE Molecular cloning of a complementary DNA encoding human macrophage-specific colony-stimulating factor (CSF-1)
JOURNAL Science 230 (4723), 291-296 (1985)
MEDLINE 86018828
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by

```
            E.S.Kawasaki, 14-JAN-1986.
            Although CSF-1 appears to be encoded by a single gene, several
            CSF-1 mRNA transcripts of different sizes are synthesized by the
            MIA-PaCa cells.  Part of the mRNA size heterogeneity can be
            accounted for by the fact that at least two of the transcripts have
            different polyadenylation signals.  Also, an inactive clone,
            pcCSF-12, as well as eight other CSF-1 cDNA clones, contain 115
            bases of intron sequence immediately preceding exon 2 of clone
            pcCSF-17.  Clone pcCSF-12 contains the entire coding sequence for
            the mature CSF-1 protein, but is inactive because there is a stop
            codon immediately after the beginning of the intron, resulting in a
            severely truncated protein.  [1] suggest that the cDNA clones were
            derived from mRNA precursors that were differently spliced.
FEATURES              Location/Qualifiers
     source           1. .361
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="5q33"
                      /cell_line="MIA-PaCa-2"
                      /tissue_type="pancreatic, tumor"
                      /tissue_type="pancreatic tumor"
     exon             <179. .340
                      /gene="CSF1"
                      /note="G00-119-811"
                      /number=1
     sig_peptide      179. .274
                      /gene="CSF1"
                      /note="G00-119-811"
     intron           341. .>361
                      /gene="CSF1"
                      /note="G00-119-811"
                      /number=1
BASE COUNT       51 a    118 c    134 g     58 t
ORIGIN      15 bp upstream of SmaI site.


  Query Match             100.0%;  Score 42;  DB 9;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT   2
E03655
LOCUS       E03655                   693 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for M-CSF active polypeptide,M1-del.
ACCESSION   E03655
VERSION     E03655.1  GI:2171870
KEYWORDS    JP 1992148687-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Mizoguchi,J., Nogawa,M., Yamashita,T., Kubota,A., Maeda,T.,
            Yanai,M., Kobayashi,J., Sato,Y. and Kotani,M. .
  TITLE     STABLY HIGHLY MANIFESTATIVE RECOMBINANT VECTOR FOR M-CSF ACTIVE
            POLYPEPTIDE
  JOURNAL   Patent: JP 1992148687-A 1 21-MAY-1992;
            ASAHI CHEM IND CO LTD
COMMENT     OS   Homo Sapiens (human)
            PN   JP 1992148687-A/1
            PD   21-MAY-1992
            PF   09-OCT-1990 JP 1990271951
            PI   MIZOGUCHI JUNZO, NOGAWA MAKOTO, YAMASHITA TOMOYUKI, PI
            KUBOTA AKIKO,
            PI   MAEDA TADASHI, YANAI MASAHIKO, KOBAYASHI JUNJI, SATO YUTAKA,
            PI   KOTANI MASARU
            PC   C12N15/85,C12N5/10,C12N15/27,C12P21/02,(C12P21/02,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: cell_line=TRC-29R(FERM BP-2375);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .693
            FT                   /product='M1-del'.
FEATURES              Location/Qualifiers
     source           1. .693
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
BASE COUNT      175 a    177 c    193 g    148 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 693;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 92


RESULT   3
I01437
LOCUS       I01437                   821 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 4 from Patent US 4847201.
ACCESSION   I01437
VERSION     I01437.1  GI:270176
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   Kaswasaki,E.S., Ladner,M.B., Van Arsdell,J.N., Wang,A.M., Ralph,P.,
            Coyne,M.Y. and Warren,M.K.
  TITLE     DNA encoding for CSF-1 and accompanying recombinant systems
  JOURNAL   Patent: US 4847201-A 4 11-JUL-1989;
            Cetus Corporation;
            Emeryville, CA
FEATURES              Location/Qualifiers
     source           1. .821
                      /organism="unknown"
BASE COUNT      167 a    204 c    251 g    199 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 821;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT   4
I09299
LOCUS       I09299                   821 bp            linear   PAT 02-DEC-1994
DEFINITION  Sequence 3 from Patent WO 8902746.
ACCESSION   I09299
VERSION     I09299.1  GI:587992
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 821)
  AUTHORS   Ralph,P., Warren,M.K., Chong,K.T., Devlin,J.J. and Zimmerman,R.
  TITLE     USES OF RECOMBINANT COLONY STIMULATING FACTOR-1
  JOURNAL   Patent: WO 8902746-A 3 06-APR-1989;
FEATURES              Location/Qualifiers
```

```
     source          1. .821
                     /organism="unknown"
BASE COUNT      167 a    204 c    251 g    199 t
ORIGIN


  Query Match           100.0%;  Score 42;  DB 6;  Length 821;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT   5
HUMCSF1X
LOCUS       HUMCSF1X                931 bp    mRNA    linear   PRI 11-FEB-2002
DEFINITION  Homo sapiens colony stimulating factor 1 (CSF1) mRNA, partial cds.
ACCESSION   M76453 S85344
VERSION     M76453.1  GI:393119
KEYWORDS    colony stimulating factor 1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Pampfer,S., Tabibzadeh,S., Chuan,F.C. and Pollard,J.W.
  TITLE     Expression of colony-stimulating factor-1 (CSF-1) messenger RNA in
            human endometrial glands during the menstrual cycle: molecular
            cloning of a novel transcript that predicts a cell surface form of
            CSF-1
  JOURNAL   Mol. Endocrinol. 5 (12), 1931-1938 (1991)
  MEDLINE   92168031
COMMENT     On Feb 8, 2002 this sequence version replaced gi:246161.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q33"
                     /tissue_type="uterine endometrial gland"
     gene            <1. .931
                     /gene="CSF1"
     CDS             <1. .735
                     /gene="CSF1"
                     /codon_start=1
                     /product="colony stimulating factor 1"
                     /protein_id="AAA52120.2"
                     /db_xref="GI:18644893"
                     /db_xref="GDB:G00-119-811"
                     /translation="TTWLGSLLLLVCLLASRSITEEVSEYCSHMIGSGHLQSLQRLID
                     SQMETSCQITFEFVDQEQLKDPVCYLKKAFLLVQYIMEDTMRFRDNTPNAIAIVQLQE
                     LSLRLKSCFTKDYEEHDKACVRTFYETPLQLLEKVKNVFNETKNLLDKDWNIFSKNCN
                     NSFAECSSQGHERQSEGSSSPQLQESVFHLLVPSVILVLLAVGGLLFYRWRRRSHQEP
                     QRADSPLEQPEGSPLTQDDRQVELPV"
BASE COUNT      221 a    254 c    256 g    200 t
ORIGIN


  Query Match           100.0%;  Score 42;  DB 9;  Length 931;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     15 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 56


RESULT   6
E12942
LOCUS       E12942                  1119 bp    DNA     linear   PAT 24-JUN-1998
DEFINITION  Human cDNA encoding M-CSF precursor.
ACCESSION   E12942
VERSION     E12942.1  GI:3251773
KEYWORDS    JP 1997107963-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Takahashi,M., Hirato,T., Nakai,S., Ko,H., Kono,N. and Hirai,Y. ..
  TITLE     PRODUCTION OF HUMAN M-CSF
  JOURNAL   Patent: JP 1997107963-A 1 28-APR-1997;
            OTSUKA PHARMACEUT CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1997107963-A/1
            PD   28-APR-1997
            PF   04-SEP-1987 JP 1996133284
            PR   17-SEP-1986 JP 86P   220750, 19-JAN-1987 JP 87P    11025, PR
            02-JUL-1987 JP 87P   166388
            PI   TAKAHASHI MASAYUKI, HIRATO TORU, NAKAI SATORU, KO HIDEMITSU,
            PI   KONO NAOMI,
            PI   HIRAI YOSHIKATSU
            PC   C12N15/00,C12P21/02//C12N5/10,(C12P21/02,C12R1:91); CC
            strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1119
            FT                   /organism='Homo sapiens'
            FT                   /cell_type='T cell cultured cell' FT
                    /cell_line='AGR-ON'
            FT                   /clone='lambda cM5'
            FT   CDS             1. .1119
            FT                   /product='M-CSF precursor'.
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      257 a    339 c    300 g    223 t
ORIGIN


  Query Match           100.0%;  Score 42;  DB 6;  Length 1119;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     51 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 92


RESULT   7
HSU22386
LOCUS       HSU22386                1370 bp    mRNA    linear   PRI 05-APR-1995
DEFINITION  Human macrophage-colony stimulating factor gamma precursor mRNA,
            complete cds.
ACCESSION   U22386
VERSION     U22386.1  GI:758780
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1370)
  AUTHORS   Cerretti,D.P., Wignall,J., Anderson,D., Tushinski,R.J.,
            Gallis,B.M., Stya,M., Gillis,S., Urdal,D.L. and Cosman,D.
  TITLE     Human macrophage-colony stimulating factor: alternative RNA and
            protein processing from a single gene
  JOURNAL   Mol. Immunol. 25 (8), 761-770 (1988)
  MEDLINE   89039923
REFERENCE   2  (bases 1 to 1370)
  AUTHORS   Cerretti,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-1995) Douglas P. Cerretti, Immunex Corporation,
```

```
          51 University St., Seattle, WA 98101, USA
FEATURES             Location/Qualifiers
     source          1. .1370
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1p13-p21"
     CDS             45. .1361
                     /note="m-csf gamma precursor"
                     /codon_start=1
                     /product="macrophage-colony stimulating factor gamma
                     precursor"
                     /protein_id="AAA64849.1"
                     /db_xref="GI:758781"
                     /translation="MTAPGAAGRCPPTTWLGSLLLLVCLLASRSITEEVSEYCSHMIG
                     SGHLQSLQRLIDSQMETSCQITFEFVDQEQLKDPVCYLKKAFLLVQYIMEDTMRFRDN
                     TPNAIAIVQLQELSLRLKSCFTKDYEEHDKACVRTFYETPLQLLEKVKNVFNETKNLL
                     DKDWNIFSKNCNNSFAECSSQDVVTKPDCNCLYPKAIPSSDPASVSPHQPLAPSMAPV
                     AGLTWEDSEGTEGSSLLPGEQPLHTVDPGSAKQRPPRSTCQSFEPPETPVVKDSTIGG
                     SPQPRPSVGAFNPGMEDILDSAMGTNWVPEEASGEASEIPVPQGTELSPSRPGGGSMQ
                     TEPARPSNFLSASSPLPASAKGQQPADVTGHERQSEGSSSPQLQESVFHLLVPSVILV
                     LLAVGGLLFYRWRRRSHQEPQRADSPLEQPEGSPLTQDDRQVELPV"
     sig_peptide     45. .140
     mat_peptide     141. .1358
                     /note="m-csf gamma"
                     /product="macrophage-colony stimulating factor gamma"
     misc_feature    1182. .1250
                     /note="encodes transmembrane domain"
BASE COUNT      305 a    416 c    382 g    267 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 9;  Length 1370;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     95 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 136


RESULT   8
I08431
LOCUS       I08431                  1640 bp            linear   PAT 02-DEC-1994
DEFINITION  Sequence 4 from Patent WO 8604607.
ACCESSION   I08431
VERSION     I08431.1  GI:588857
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1640)
  AUTHORS   Kawasaki,E.S., Ladner,M.B., Van Arsdell,J.N., Wang,A.M. and
            Ralph,P.
  TITLE     RECOMBINANT COLONY STIMULATING FACTOR-1
  JOURNAL   Patent: WO 8604607-A 4 14-AUG-1986;
FEATURES             Location/Qualifiers
     source          1. .1640
                     /organism="unknown"
BASE COUNT      356 a    480 c    402 g    402 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1640;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    228 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 269


RESULT   9
I00368
LOCUS       I00368                  1641 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 1 from Patent US 4929700.
ACCESSION   I00368
VERSION     I00368.1  GI:314019
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1641)
  AUTHORS   Halenbeck,R., Koths,K., Cowgill,C. and Laird,W.J.
  TITLE     Production of purified, biologically active, bacterially produced
            recombinant human CSF-1
  JOURNAL   Patent: US 4929700-A 1 29-MAY-1990;
            Cetus Corporation;
            Emeryville, CA
COMMENT     On Jul 30, 1993 this sequence version replaced gi:285550.
FEATURES             Location/Qualifiers
     source          1. .1641
                     /organism="unknown"
BASE COUNT      356 a    476 c    406 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1641;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  10
I01429
LOCUS       I01429                  1641 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 1 from Patent US 4847325.
ACCESSION   I01429
VERSION     I01429.1  GI:270181
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1641)
  AUTHORS   Shadle,P.J., Koths,K.E., Moreland,M. and Katre,N.
  TITLE     Conjugation of polymer to colony stimulating factor-1
  JOURNAL   Patent: US 4847325-A 1 11-JUL-1989;
            Cetus Corporation;
            Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .1641
                     /organism="unknown"
BASE COUNT      356 a    479 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1641;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  11
AR105651
LOCUS       AR105651                1642 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 3 from patent US 6103224.
ACCESSION   AR105651
VERSION     AR105651.1  GI:12819716
KEYWORDS    .
SOURCE      Unknown.
```

```
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Ladner,M.B., Van Arsdell,J.N., Martin,G.A., Kawasaki,E.S.,
            Coyne,M.Yee, Halenbeck,R.F. and Koths,K.E.
  TITLE     N.gradient.2 CSF-1 (short form) and carboxy truncated fragments
            thereof
  JOURNAL   Patent: US 6103224-A 3 15-AUG-2000;
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="unknown"
BASE COUNT      356 a    480 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  12
AR110609
LOCUS       AR110609                1642 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 3 from patent US 6117422.
ACCESSION   AR110609
VERSION     AR110609.1  GI:12827423
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Ladner,M.B., Noble,J.A., Martin,G.A., Kawasaki,E.S., Coyne,M.Yee,
            Halenbeck,R.F. and Koths,K.E.
  TITLE     N.gradient.2-CSF-1(long form) and carboxy truncated fragments
            thereof
  JOURNAL   Patent: US 6117422-A 3 12-SEP-2000;
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="unknown"
BASE COUNT      356 a    480 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  13
AR141545
LOCUS       AR141545                1642 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6146851.
ACCESSION   AR141545
VERSION     AR141545.1  GI:15101061
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Ladner,M.B., Van Arsdell,J.N., Martin,G.A., Kawasaki,E.S.,
            Coyne,M.Yee, Halenbeck,R.F. and Koths,K.E.
  TITLE     DNA encoding NV2 (long form) and carboxy truncated fragments
            thereof
  JOURNAL   Patent: US 6146851-A 3 14-NOV-2000;
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="unknown"
BASE COUNT      356 a    480 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  14
AR142870
LOCUS       AR142870                1642 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6204020.
ACCESSION   AR142870
VERSION     AR142870.1  GI:15104156
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Ladner,M.B., Van Arsdell,J.N., Martin,G.A., Kawasaki,E.S.,
            Coyne,M.Yee, Halenbeck,R.F. and Koths,K.E.
  TITLE     DNA encoding N.gradient.2 CSF-1 (short form) and carboxy truncated
            fragment thereof
  JOURNAL   Patent: US 6204020-A 3 20-MAR-2001;
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="unknown"
BASE COUNT      356 a    480 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db    229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270


RESULT  15
I01440
LOCUS       I01440                  1642 bp ss-DNA     linear   PAT 21-MAY-1993
DEFINITION  Sequence 7 from Patent US 4847201.
ACCESSION   I01440
VERSION     I01440.1  GI:270179
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1642)
  AUTHORS   Kaswasaki,E.S., Ladner,M.B., Van Arsdell,J.N., Wang,A.M., Ralph,P.,
            Coyne,M.Y. and Warren,M.K.
  TITLE     DNA encoding for CSF-1 and accompanying recombinant systems
  JOURNAL   Patent: US 4847201-A 7 11-JUL-1989;
            Cetus Corporation;
            Emeryville, CA
FEATURES             Location/Qualifiers
     source          1. .1642
                     /organism="unknown"
BASE COUNT      356 a    480 c    403 g    403 t
ORIGIN


  Query Match             100.0%;  Score 42;  DB 6;  Length 1642;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
```

Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
           |||||||||||||||||||||||||||||||||||||||||||
Db     229 CTCCCTGCTGTTGTTGGTCTGTCTCCTGGCGAGCAGGAGTAT 270
```

Search completed: September 11, 2002, 04:49:11
Job time: 5923 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 11, 2002, 03:52:28 ; Search time 177.6 Seconds
(without alignments)
406.027 Million cell updates/sec

Title: US-09-786-214-11
Perfect score: 42
Sequence: 1 ctccctgctgttgttggtct.....ctcctggcgagcaggagtat 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	100.0	42	21	AAZ99677	DNA derived from m
2	42	100.0	60	21	AAZ99675	Truncated alternat
3	42	100.0	78	21	AAZ99672	Alternate reading
4	42	100.0	129	21	AAZ99758	Partial major macr
5	42	100.0	660	20	AAX80211	Human CSF-1 nucleo
6	42	100.0	693	13	AAQ24188	M1-del gene. Homo
7	42	100.0	756	19	AAV33210	Macrophage colony
8	42	100.0	771	14	AAQ38875	Sequence of the M-
9	42	100.0	771	19	AAV33209	Macrophage colony
10	42	100.0	861	7	AAN60588	First half of part
11	42	100.0	1119	10	AAN90515	Human macrophage c
12	42	100.0	1317	14	AAQ38876	Sequence of the M-
13	42	100.0	1640	7	AAN60590	Sequence encoding
14	42	100.0	1641	11	AAQ04901	Clone pcCSF17 enco
15	42	100.0	1641	17	AAT08561	Human CSF-1 cDNA c
16	42	100.0	1642	9	AAN80775	Sequence of a cDNA
17	42	100.0	1642	10	AAN90499	cDNA of plasmid pC
18	42	100.0	1642	10	AAN90748	Sequence encoding
19	42	100.0	1642	10	AAN91084	Colony stimulating
20	42	100.0	1642	17	AAT58753	Human short form C
21	42	100.0	1642	18	AAV01080	Gene for human col
22	42	100.0	1642	18	AAT94748	Truncated human co
23	42	100.0	1642	18	AAT80591	Human colony stimu
24	42	100.0	1642	20	AAX23992	Human pcCSF-17 cDN
25	42	100.0	1642	21	AAA88466	Human colony stimu
26	42	100.0	1642	21	AAA37721	Human CSF-1 short
27	42	100.0	1642	22	AAD03860	Human colony stimu
28	42	100.0	1642	22	AAC93061	Human colony stimu
29	42	100.0	1642	22	AAC89098	Human SCSF-1 codin
30	42	100.0	1665	14	AAQ33028	Sequence of the M-
31	42	100.0	1845	10	AAN90601	HindIII fragment o
32	42	100.0	1846	17	AAT08562	Human CSF-1 genomi
33	42	100.0	2045	10	AAN90749	Human colony stimu
34	42	100.0	2237	15	AAQ56772	Human MCSF cDNA.
35	42	100.0	2488	10	AAN90514	cDNA of lambda-cM5
36	42	100.0	2488	14	AAQ43709	Sequence of clone
37	42	100.0	2490	9	AAN81382	Sequence encoding
38	42	100.0	2500	9	AAN81384	Sequence encoding
39	42	100.0	2500	10	AAN90516	Lambda-cM11 cDNA e
40	42	100.0	2500	14	AAQ43710	Sequence of clone
41	42	100.0	3391	24	AAS94797	Human DNA sequence
42	42	100.0	3981	8	AAN70695	Macrophage-colony
43	42	100.0	3981	10	AAN91734	Sequence of a macr
44	42	100.0	4021	12	AAQ10343	Monocyte macropha
45	40	95.2	2037	10	AAN91085	Clone 4 contg. pcC

ALIGNMENTS

RESULT 1
AAZ99677
ID AAZ99677 standard; DNA; 42 BP.
XX
AC AAZ99677;
XX
DT 12-JUL-2000 (first entry)
XX
DE DNA derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..42
FT /*tag= a
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

```
XX
DR   WPI; 2000-256859/22.
DR   P-PSDB; AAY84266.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame
PT   macrophage-colony stimulating factor comprises 25 amino acid residue
PT   sequence -
XX
PS   Example 2; Page 65; 74pp; English.
XX
CC   The present sequence encodes a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics.
XX
SQ   Sequence 42 BP; 4 A; 11 C; 13 G; 14 T; 0 other;


  Query Match             100.0%;  Score 42;  DB 21;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42


RESULT   2
AAZ99675
ID   AAZ99675 standard; DNA; 60 BP.
XX
AC   AAZ99675;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Truncated alternate reading frame of macrophage colony stimulating gene.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..60
FT                   /*tag=  a
FT                   /product= "tumour rejection antigen"
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US20344.
XX
PR   04-SEP-1998;   98US-0099077.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   P-PSDB; AAY85265.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame
PT   macrophage-colony stimulating factor comprises 25 amino acid residue
PT   sequence -
XX
PS   Example 2; Page 64; 74pp; English.
XX
CC   The present sequence encodes a truncated tumour rejection antigen
CC   precursor, and represents a truncated alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics.
XX
SQ   Sequence 60 BP; 8 A; 15 C; 21 G; 16 T; 0 other;


  Query Match             100.0%;  Score 42;  DB 21;  Length 60;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51


RESULT   3
AAZ99672
ID   AAZ99672 standard; DNA; 78 BP.
XX
AC   AAZ99672;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Alternate reading frame of macrophage colony stimulating gene.
XX
KW   Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..78
FT                   /*tag=  a
FT                   /product= "tumour rejection antigen"
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US20344.
XX
PR   04-SEP-1998;   98US-0099077.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
DR   P-PSDB; AAY84264.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame
PT   macrophage-colony stimulating factor comprises 25 amino acid residue
PT   sequence -
XX
```

PS Example 2; Page 63; 74pp; English.
XX
CC The present sequence encodes a tumour rejection antigen precursor,
CC and represents an alternative open reading frame (ORF) of human
CC macrophage colony stimulating gene. Peptides derived from the
CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 78 BP; 11 A; 17 C; 29 G; 21 T; 0 other;

Query Match 100.0%; Score 42; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
||||||||||||||||||||||||||||||||||||||||||
Db 10 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 51

RESULT 4
AAZ99758
ID AAZ99758 standard; DNA; 129 BP.
XX
AC AAZ99758;
XX
DT 12-JUL-2000 (first entry)
XX
DE Partial major macrophage colony stimulating gene sequence.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..129
FT /*tag= a
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR P-PSDB; AAY69496.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Disclosure; Fig 4; 74pp; English.
XX
CC The present sequence encodes a portion of macrophage-colony stimulating
CC factor. The specification describes a tumour rejection antigen precursor,
CC which is encoded by an alternative open reading frame (ORF) of human
CC macrophage colony stimulating gene. Peptides derived from the
CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 129 BP; 18 A; 40 C; 44 G; 27 T; 0 other;

Query Match 100.0%; Score 42; DB 21; Length 129;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
||||||||||||||||||||||||||||||||||||||||||
Db 51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92

RESULT 5
AAX80211
ID AAX80211 standard; DNA; 660 BP.
XX
AC AAX80211;
XX
DT 20-AUG-1999 (first entry)
XX
DE Human CSF-1 nucleotide sequence.
XX
KW Human; M-CSF-1; colony stimulating factor; angiogenesis; tumour;
KW proliferative disorder; wound healing; cellular proliferative disease;
KW diabetic retinopathy; cancer; solid tumour; ss.
XX
OS Homo sapiens.
XX
PN WO9929345-A1.
XX
PD 17-JUN-1999.
XX
PF 04-DEC-1998; 98WO-US25791.
XX
PR 05-DEC-1997; 97US-0067591.
XX
PA (LJOL-) LA JOLLA INST EXPERIMENTAL MEDICINE.
XX
PI Borgstrom P, Bourdon MA, Deryugina E, Rao PS;
XX
DR WPI; 1999-385494/32.
XX
PT Inhibition of angiogenesis by macrophage intervention, useful for
PT inhibiting tumour growth
XX
PS Disclosure; Page 28; 30pp; English.
XX
CC The present invention describes a method for the inhibition a
CC angiogenesis in a cell population in a mammal comprising inhibiting a
CC host cells angiogenic effect in the said mammal. The method can be
CC used to inhibit angiogenesis in tumours, wounds surrounding cells or
CC cells characteristic of a proliferative disorder. Therefore the method
CC is useful in the treatment of wound healing and cellular proliferative
CC diseases, e.g. diabetic retinopathy, and cancers, especially associated
CC with the presence of solid tumours. The method can be used in combination
CC with additional treatment, including surgery, radiation therapy and
CC chemotherapy. The present sequence represents a nucleotide sequence of
CC human colony stimulating factor (CSF-1) as given in the present
CC invention.
XX
SQ Sequence 660 BP; 154 A; 186 C; 182 G; 138 T; 0 other;

```
  Query Match            100.0%;  Score 42;  DB 20;  Length 660;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    156 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 197


RESULT   6
AAQ24188
ID   AAQ24188 standard; DNA; 693 BP.
XX
AC   AAQ24188;
XX
DT   23-OCT-1992  (first entry)
XX
DE   M1-del gene.
XX
KW   M-CSF delete DNA; dhfr; SV40; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     97..690
FT                   /*tag=  a
FT   sig_peptide     1..96
FT                   /*tag=  b
XX
PN   GB2249100-A.
XX
PD   29-APR-1992.
XX
PF   04-OCT-1991;   91GB-0021220.
XX
PR   09-OCT-1990;   90JP-0271951.
XX
PA   (TOXN ) TOYO JOZO KK.
XX
PI   Kobayashi J,  Kubota A,  Maeda T,  Mizoguchi J,  Nogawa M;
PI   Otani M,  Satoh S,  Taniuchi M,  Yamashita T;
XX
DR   WPI; 1992-143520/18.
DR   P-PSDB; AAR23649.
XX
PT   Recombinant vector contg. M-CSF-delete DNA -  used to produce
PT   transformants with high level of M-CSF activity
XX
PS   Disclosure; Page 34; 48pp; English.
XX
CC   The sequence given is the nucleotide sequence of M-CSF-delete DNA,
CC   also referred to as the M1-del gene.
CC   A recombinant vector containing M-CSF-delete DNA also contains
CC   dehydrofolate reductase gene (dhfr) marker DNA both of which are
CC   located downstream of SV40 early gene promoter with a distance of
CC   1.5 kb.  The recombinant vector allows stable prodn. in a high
CC   conc., of active M-CSF.  M-CSF can be useful for research and in
CC   clinical eveluation.
XX
SQ   Sequence 693 BP; 174 A; 177 C; 194 G; 148 T; 0 other;


  Query Match            100.0%;  Score 42;  DB 13;  Length 693;
  Best Local Similarity  100.0%;  Pred. No. 3.5e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT   7
AAV33210
ID   AAV33210 standard; DNA; 756 BP.
XX
AC   AAV33210;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Macrophage colony stimulating factor-alpha (del161-165) mutant DNA.
XX
KW   Macrophage colony stimulating factor-alpha; M-CSF; human;
KW   gene therapy; cancer; Leishmania; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9839449-A1.
XX
PD   11-SEP-1998.
XX
PF   04-MAR-1998;   98WO-US04802.
XX
PR   04-MAR-1997;   97US-0038583.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Dwarki V,  Koths KE,  Manning WC;
XX
DR   WPI; 1998-495850/42.
DR   P-PSDB; AAW70329.
XX
PT   Treatment of diseased cells, e.g. cancer cells - by transforming
PT   cells with gene delivery vehicle capable of expressing a macrophage
PT   colony stimulating factor-alpha mutant, used for, e.g. treating
PT   cancer
XX
PS   Example 5; Page 64; 78pp; English.
XX
CC   This DNA sequence codes for a macrophage-colony stimulating factor
CC   alpha (M-CSFa) mutant (see AAW70329) in which amino acid residues
CC   161-165 of wild-type G-CSFa (see AAW70328) are deleted.  It was
CC   obtained by site-directed mutagenesis of G-CSFa cDNA.  A claimed
CC   method for reducing a population of diseased cells comprises
CC   transfecting or transducing at least 1 from the population of
CC   diseased cells with a gene delivery vehicle (GDV) capable of
CC   expressing a M-CSFa mutant.  Such mutants preferably have a
CC   deletion or substitution of amino acids 156-160, 159-165, 161-165,
CC   161-162 or 163-165.  They have a decreased capacity to be
CC   proteolytically processed and released from a cell membrane and
CC   thus remain on the membrane longer than the wild-type protein.
CC   Also claimed is a method for reducing a population of diseased
CC   cells by: (i) administering a GDV which infects a population of
CC   diseased cells and which expresses an M-CSFa polypeptide and a
CC   prodrug activator polypeptide, and (ii) administering a prodrug
CC   which is converted by the prodrug activator polypeptide in the
CC   diseased cells to an agent that is toxic to the diseased cells,
CC   where at least some cells in the population of diseased cells are
CC   destroyed.  Expression of the M-CSFa mutant in the population of
CC   cells can provide an animal with an increased immune response
CC   against the diseased cells.  The  methods can be used for treating
CC   diseases such as cancer or a disease manifesting a population of
CC   aberrant cells where the population is created by infection with a
CC   pathogen, e.g. Leishmania.
XX
SQ   Sequence 756 BP; 180 A; 199 C; 211 G; 166 T; 0 other;


  Query Match            100.0%;  Score 42;  DB 19;  Length 756;
  Best Local Similarity  100.0%;  Pred. No. 3.6e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92
```

```
RESULT   8
AAQ38875
ID   AAQ38875 standard; cDNA; 771 BP.
XX
AC   AAQ38875;
XX
DT   06-MAY-1993  (first entry)
XX
DE   Sequence of the M-CSF clone, M-CSF -alpha.
XX
KW   Macrophage; colony stimulating factor; immune system; proliferation;
KW   differentiation; interferon-gamma; TNF; splicing; ss.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..771
FT                   /*tag= a
FT                   /note= "encodes M-CSF-alpha"
FT   sig_peptide     1..96
FT                   /*tag= b
FT   mat_peptide     97..771
FT                   /*tag= c
XX
PN   US5171675-A.
XX
PD   15-DEC-1992.
XX
PF   28-JUL-1988;   88US-0226050.
XX
PR   28-JUL-1988;   88US-0226050.
PR   08-AUG-1989;   89US-0391218.
XX
PA   (CERR/) CERRETTI D P.
XX
PI   Anderson DM,  Cerretti DP,  Cosman D,  Gallis BM,  Tushinski RJ;
XX
DR   WPI; 1993-008579/01.
DR   P-PSDB; AAR30197.
XX
PT   Isolated DNA useful for determining biological role of M-CSF -
PT   comprises nucleotide sequence encoding a functional biologically
PT   active human macrophage colony stimulating factor gamma protein
XX
PS   Disclosure; Fig 2; 18pp; English.
XX
CC   The sequence represents M-CSF clone -alpha, which is a primary
CC   translation prod. of an alternative splicing event of M-CSF, and
CC   lacks bases 545 to 1438 of M-CSF -beta.  The sequence represents
CC   a precursor to biologically active M-CSF which is important in the
CC   regulation of immune responses, by its ability to potentiate the
CC   proliferation and differentiation of macrophages from immature
CC   haematopoietic progenitor cells, and induces effector functions of
CC   mature macrophages including secretion of interferon gamma, TNF and
CC   non-M-CSF colony stimulating activities.
CC   See also AAQ33028 and AAQ38876.
XX
SQ   Sequence 771 BP; 183 A; 204 C; 216 G; 168 T; 0 other;

  Query Match             100.0%;  Score 42;  DB 14;  Length 771;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT   9
AAV33209
ID   AAV33209 standard; DNA; 771 BP.
XX
AC   AAV33209;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Macrophage colony stimulating factor-alpha DNA.
XX
KW   Macrophage colony stimulating factor-alpha; M-CSF; human;
KW   gene therapy; cancer; Leishmania; ss.
XX
OS   Homo sapiens.
XX
PN   WO9839449-A1.
XX
PD   11-SEP-1998.
XX
PF   04-MAR-1998;   98WO-US04802.
XX
PR   04-MAR-1997;   97US-0038583.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Dwarki V,  Koths KE,  Manning WC;
XX
DR   WPI; 1998-495850/42.
DR   P-PSDB; AAW70328.
XX
PT   Treatment of diseased cells, e.g. cancer cells - by transforming
PT   cells with gene delivery vehicle capable of expressing a macrophage
PT   colony stimulating factor-alpha mutant, used for, e.g. treating
PT   cancer
XX
PS   Example 7; Page 62-63; 78pp; English.
XX
CC   This DNA sequence codes for macrophage-colony stimulating factor
CC   alpha (M-CSFa, see AAW70328).  A claimed method for reducing a
CC   population of diseased cells comprises transfecting or transducing
CC   at least 1 from the population of diseased cells with a gene
CC   delivery vehicle (GDV) capable of expressing a M-CSFa mutant.  Such
CC   mutants (see AAW70329) have a decreased capacity to be proteolytically
CC   processed and released from a cell membrane.  Also claimed is a
CC   method for reducing a population of diseased cells by: (i)
CC   administering a GDV which infects a population of diseased cells
CC   and which expresses an M-CSFa polypeptide and a prodrug activator
CC   polypeptide, and (ii) administering a prodrug which is converted by
CC   the prodrug activator polypeptide in the diseased cells to an agent
CC   that is toxic to the diseased cells, where at least some cells in
CC   the population of diseased cells are destroyed.  Expression of the
CC   M-CSFa mutant in the population of cells can provide an animal with
CC   an increased immune response against the diseased cells.  The
CC   methods can be used for treating diseases such as cancer or a
CC   disease manifesting a population of aberrant cells where the
CC   population is created by infection with a pathogen, e.g. Leishmania.
XX
SQ   Sequence 771 BP; 183 A; 205 C; 216 G; 167 T; 0 other;

  Query Match             100.0%;  Score 42;  DB 19;  Length 771;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          |||||||||||||||||||||||||||||||||||||||||||
Db     51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT  10
AAN60588
ID   AAN60588 standard; cDNA; 861 BP.
XX
AC   AAN60588;
XX
DT   29-JUL-1991  (first entry)
XX
DE   First half of partial sequence encoding human colony stimulating
```

```
DE   factor (CSF).
XX
KW   Immune system; enhancement; interferon production; monocyte; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            218..340
FT                   /*tag=  a
XX
PN   WO8604607-A.
XX
PD   14-AUG-1986.
XX
PF   03-FEB-1986;    86WO-US00238.
XX
PR   21-JAN-1986;    86US-0821068.
PR   05-FEB-1985;    85US-0698359.
PR   30-APR-1985;    85US-0728834.
PR   14-JUN-1985;    85US-0744924.
PR   18-JUL-1985;    85US-0756814.
PR   20-JUN-1986;    86US-0876819.
XX
PA   (CETU ) CETUS CORP.
XX
PI   Kawasaki ES,  Ladner MB,  VanArsdell JN,  Wang AM,  Ralph P;
PI   Coyne MY,  Warren MK;
XX
DR   WPI; 1986-225459/34.
DR   P-PSDB; AAP61505.
XX
PT   New recombinant colony stimulating factor-1 protein - is obtd. by
PT   recombinant DNA procedures for use in enhancing immune system and
PT   for treating monocytes to increase interferon prodn.
XX
PS   Disclosure; Fig 4a-b; 82pp; English.
XX
CC   The CSF-1 protein claimed is esp. human CSF-1, or it may have an
CC   equiv. AA sequence, with deletion or conservative substitution of
CC   one or more AA, esp. between positions 158 and 224; 51 and 52 and/or
CC   191,192 and193; 15-20 and/or 75-84. The protein is esp. mCSF-1, or
CC   158 deleted CSF-1 or asp (59)-CSF1.
XX
SQ   Sequence 861 BP; 167 A; 204 C; 251 G; 199 T; 40 other;


  Query Match             100.0%;  Score 42;  DB 7;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
           |||||||||||||||||||||||||||||||||||||||||||
Db     229 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 270


RESULT  11
AAN90515
ID   AAN90515 standard; cDNA; 1119 BP.
XX
AC   AAN90515;
XX
DT   27-NOV-1989  (first entry)
XX
DE   Human macrophage colony stimulating factor precursor cDNA.
XX
KW   Macrophage colony stimulating factor; human; recombinant;
KW   leukopoenia; bone marrow transplant; anti-cancer drug;
KW   infectious diseases; AGR-ON cell line; human leukaemic T cells;
KW   precursor protein.
XX
OS   Homo sapiens (human).
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1116
FT                   /*tag=  a
XX
PN   EP328061-A.
XX
PD   16-AUG-1989.
XX
PF   08-FEB-1989;    89EP-0102138.
XX
PR   08-FEB-1988;    88JP-0062841.
XX
PA   (SAKA ) OTSUKA PHARM KK.
XX
PI   Hirato T,  Nakai S,  Hong YM,  Kouno N,  Hirai Y;
XX
DR   WPI; 1989-235429/33.
DR   P-PSDB; AAP91031.
XX
PT   Human recombinant colony-stimulating factors
PT   - obtd. using AGR-ON cell line, used for treating leukopoenia
PT   and against infectious diseases.
XX
PS   Disclosure; fig 3; 73pp; English.
XX
CC   cDNA encoding biologically active recombinant human macrophage colony
CC   stimulating factor precusor protein (AAP91031). This can be used
CC   as a drug for prevention and cure of diseases involving leukopoenia,
CC   as an auxillary agent for bone marrow transplants, as a drug to
CC   prevent and cure infectious diseases, and as an anti-cancer drug.
CC   See AAN90514-N90522, AAP91032-P91043, AAP90123.
XX
SQ   Sequence 1119 BP; 257 A; 339 C; 300 G; 223 T; 0 other;


  Query Match             100.0%;  Score 42;  DB 10;  Length 1119;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
           |||||||||||||||||||||||||||||||||||||||||||
Db      51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92


RESULT  12
AAQ38876
ID   AAQ38876 standard; cDNA; 1317 BP.
XX
AC   AAQ38876;
XX
DT   06-MAY-1993  (first entry)
XX
DE   Sequence of the M-CSF clone, M-CSF -gamma.
XX
KW   Macrophage; colony stimulating factor; immune system; proliferation;
KW   differentiation; interferon-gamma; TNF; splicing; ss.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1317
FT                   /*tag=  a
FT                   /note= "encodes M-CSF-gamma"
FT   sig_peptide     1..96
FT                   /*tag=  b
FT   mat_peptide     97..1317
FT                   /*tag=  c
XX
PN   US5171675-A.
XX
PD   15-DEC-1992.
XX
PF   28-JUL-1988;    88US-0226050.
XX
PR   28-JUL-1988;    88US-0226050.
PR   08-AUG-1989;    89US-0391218.
```

XX
PA (CERR/) CERRETTI D P.
XX
PI Anderson DM, Cerretti DP, Cosman D, Gallis BM, Tushinski RJ;
XX
DR WPI; 1993-008579/01.
DR P-PSDB; AAR30198.
XX
PT Isolated DNA useful for determining biological role of M-CSF -
PT comprises nucleotide sequence encoding a functional biologically
PT active human macrophage colony stimulating factor gamma protein
XX
PS Disclosure; Fig 2; 18pp; English.
XX
CC The sequence represents M-CSF clone -gamma, which is a primary
CC translation prod. of an alternative splicing event of M-CSF, and
CC lacks bases 1091 to 1438 of M-CSF -beta. The sequence represents
CC a precursor to biologically active M-CSF which is important in the
CC regulation of immune responses, by its ability to potentiate the
CC proliferation and differentiation of macrophages from immature
CC haematopoietic progenitor cells, and induces effector functions of
CC mature macrophages including secretion of interferon gamma, TNF and
CC non-M-CSF colony stimulating activities.
CC See also AAQ33028 and AAQ38875.
XX
SQ Sequence 1317 BP; 297 A; 395 C; 362 G; 263 T; 0 other;

Query Match 100.0%; Score 42; DB 14; Length 1317;
Best Local Similarity 100.0%; Pred. No. 4e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
||||||||||||||||||||||||||||||||||||||||||
Db 51 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 92

RESULT 13
AAN60590
ID AAN60590 standard; cDNA; 1640 BP.
XX
AC AAN60590;
XX
DT 29-JUL-1991 (first entry)
XX
DE Sequence encoding mature colony stimulating factor 1 (mCSF-1).
XX
KW Immune system; enhancement; interferon production; monocyte; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 178..948
FT /*tag= a
XX
PN WO8604607-A.
XX
PD 14-AUG-1986.
XX
PF 03-FEB-1986; 86WO-US00238.
XX
PR 21-JAN-1986; 86US-0821068.
PR 05-FEB-1985; 85US-0698359.
PR 30-APR-1985; 85US-0728834.
PR 14-JUN-1985; 85US-0744924.
PR 18-JUL-1985; 85US-0756814.
PR 20-JUN-1986; 86US-0876819.
XX
PA (CETU) CETUS CORP.
XX
PI Kawasaki ES, Ladner MB, VanArsdell JN, Wang AM, Ralph P;
PI Coyne MY, Warren MK;
XX
DR WPI; 1986-225459/34.
DR P-PSDB; AAP60667.
XX
PT New recombinant colony stimulating factor-1 protein - is obtd. by
PT recombinant DNA procedures for use in enhancing immune system and
PT for treating monocytes to increase interferon prodn.
XX
PS Disclosure; Fig 5; 82pp; English.
XX
CC The CSF-1 protein claimed is esp. human CSF-1, or it may have an
CC equiv. AA sequence, with deletion or conservative substitution of
CC one or more AA, esp. between positions 158 and 224; 51 and 52 and/or
CC 191,192 and193; 15-20 and/or 75-84. The protein is esp. mCSF-1, or
CC 158 deleted CSF-1 or asp (59)-CSF1. AAN60590 encodes a 32 residue
CC putative signal sequence, which is presumably cleaved upon
CC secretion from mammalian cells.
XX
SQ Sequence 1640 BP; 356 A; 479 C; 403 G; 402 T; 0 other;

Query Match 100.0%; Score 42; DB 7; Length 1640;
Best Local Similarity 100.0%; Pred. No. 4.1e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
||||||||||||||||||||||||||||||||||||||||||
Db 228 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 269

RESULT 14
AAQ04901
ID AAQ04901 standard; cDNA; 1641 BP.
XX
AC AAQ04901;
XX
DT 17-OCT-1990 (first entry)
XX
DE Clone pcCSF17 encoding a short form of human CSF1.
XX
KW Colony stimulating factor; CSF-1; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 179..946
FT /*tag= a
XX
PN US4929700-A.
XX
PD 29-MAY-1990.
XX
PF 08-APR-1988; 88US-0173428.
XX
PR 16-APR-1987; 87US-0040174.
PR 08-APR-1988; 88US-0173428.
XX
PA (CETU) CETUS CORP.
XX
PI Halenbeck R, Koths K, Cowgill C, Laird WJ;
XX
DR WPI; 1990-192981/25.
DR P-PSDB; AAR05486.
XX
PT Purified, biologically active colony stimulating factor-1 dimer -
PT from bacterially produced recombinant CSF-1 monomer by
PT solubilisation and refolding.
XX
PS Disclosure; ; pp; English.
XX
CC DNA useful in a method of producing active CSF-1 heterodimers from
CC the monomeric gene product, by unfolding, solubilising and
CC refolding the protein.
XX

```
SQ   Sequence 1641 BP; 356 A; 475 C; 407 G; 403 T; 0 other;

  Query Match             100.0%;  Score 42;  DB 11;  Length 1641;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    229 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 270


RESULT  15
AAT08561
ID   AAT08561 standard; cDNA; 1641 BP.
XX
AC   AAT08561;
XX
DT   27-MAR-1996  (first entry)
XX
DE   Human CSF-1 cDNA clone CSF-17.
XX
KW   CSF-1; macrophage colony stimulating factor-1; lymphokine;
KW   cytokine; immunosuppression; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             178..948
FT                   /*tag=  a
FT   sig_peptide     178..273
FT                   /*tag=  b
FT   mat_peptide     274..945
FT                   /*tag=  c
XX
PN   US5470569-A.
XX
PD   28-NOV-1995.
XX
PF   05-FEB-1985;   85US-0698359.
XX
PR   09-FEB-1988;   88US-0157094.
PR   05-FEB-1985;   85US-0698359.
PR   30-APR-1985;   85US-0728834.
PR   14-JUN-1985;   85US-0744924.
PR   18-JUL-1985;   85US-0756814.
PR   21-JAN-1986;   86US-0821068.
PR   26-MAY-1989;   89US-0358394.
PR   18-NOV-1991;   91US-0794822.
PR   14-MAR-1994;   94US-0212300.
XX
PA   (CETU ) CETUS ONCOLOGY CORP.
XX
PI   Coyne MY,  Kawasaki ES,  Ladner MB,  Ralph P,  Van Arsdell JN;
PI   Wang AM, Warren MK;
XX
DR   WPI; 1996-019830/02.
DR   P-PSDB; AAR77650.
XX
PT   Recombinant human colony-stimulating factor 1 fragments - useful for
PT   overcoming immunosuppression, e.g. caused by cancer chemotherapy.
XX
PS   Disclosure; Fig 5A-B; 29pp; English.
XX
CC   A cDNA clone (AAT08561), designated CSF-17 (CMCC 2347, ATCC 53149),
CC   codes for human colony stimulating factor-1 (CSF-1) (AAR77650).  It
CC   was obtd. from a cDNA library derived from human pancreatic
CC   carcinoma line MIAPaCa cells.  Expression of the cDNA allows
CC   prodn. of mature CSF-1 protein in procaryotic and eucaryotic
CC   host cells, and prodn of CSF-1 muteins.
XX
SQ   Sequence 1641 BP; 356 A; 478 C; 404 G; 403 T; 0 other;

  Query Match             100.0%;  Score 42;  DB 17;  Length 1641;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-06;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 42
          ||||||||||||||||||||||||||||||||||||||||||
Db    228 ctccctgctgttgttggtctgtctcctggcgagcaggagtat 269


Search completed: September 11, 2002, 04:53:13
Job time: 3645 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:50:27 ; Search time 28.7 Seconds
(without alignments)
18.888 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	59.7	211	1	COBL_METJA	Q58917 methanococc
2	41	56.9	508	1	COBI_MYCTU	Q10677 mycobacteri
3	39	54.2	115	1	KV3I_HUMAN	P04433 homo sapien
4	39	54.2	115	1	KV5I_MOUSE	P01642 mus musculu
5	39	54.2	278	1	PURU_HAEIN	Q03432 haemophilus
6	39	54.2	516	1	C4AD_DROME	Q9v4t3 drosophila
7	39	54.2	597	1	NR41_RAT	P22829 rattus norv
8	39	54.2	1402	1	Y197_MOUSE	Q9z0w3 mus musculu
9	39	54.2	1636	1	BUD3_YEAST	P25558 saccharomyc
10	38	52.8	390	1	COBL_MYCTU	Q10671 mycobacteri
11	38	52.8	429	1	RS1_LEULA	P50889 leuconostoc
12	38	52.8	507	1	CATA_PICAN	P30263 pichia angu
13	38	52.8	699	1	EFG_HAEIN	P43925 haemophilus
14	38	52.8	700	1	EFG_PASMU	P57938 pasteurella
15	37	51.4	280	1	PMXA_MOUSE	Q62066 mus musculu
16	37	51.4	281	1	PMXA_RAT	Q62782 rattus norv
17	37	51.4	446	1	COBJ_ARCFU	O29534 a cobalamin
18	37	51.4	446	1	ENO1_MAIZE	P26301 zea mays (m
19	37	51.4	557	1	PUR6_VIGAC	P55195 vigna aconi
20	37	51.4	637	1	MUTL_CAUCR	Q9rp66 caulobacter
21	37	51.4	658	1	VG18_BPT4	P13332 bacteriopha
22	37	51.4	753	1	CKAA_BACUF	O32321 bacillus th
23	37	51.4	992	1	EVC_HUMAN	P57679 homo sapien
24	37	51.4	2269	1	RRPL_SV41	P35341 simian viru
25	36.5	50.7	374	1	RGSK_BOVIN	P79348 bos taurus
26	36	50.0	129	1	KV3H_HUMAN	P04207 homo sapien
27	36	50.0	158	1	RSD_ECOLI	P31690 escherichia
28	36	50.0	162	1	RSD_SALTY	Q9l9i6 salmonella
29	36	50.0	243	1	SUMT_SYNP7	P42451 synechococc
30	36	50.0	280	1	PURU_ECOLI	P37051 escherichia
31	36	50.0	280	1	PURU_SHIFL	P38480 shigella fl
32	36	50.0	342	1	HUPK_AZOVI	P31878 azotobacter
33	36	50.0	348	1	HOXV_AZOVI	P40597 azotobacter
34	36	50.0	359	1	ALF2_PEA	P46257 pisum sativ
35	36	50.0	359	1	ALF_CICAR	O65735 cicer ariet
36	36	50.0	422	1	TKSU_PYRKO	P58502 pyrococcus
37	36	50.0	471	1	CD36_BOVIN	P26201 bos taurus
38	36	50.0	598	1	NR41_CANFA	P51666 canis famil
39	36	50.0	750	1	ELS_CHICK	P07916 gallus gall
40	36	50.0	844	1	HEXA_STRPN	P10564 streptococc
41	36	50.0	926	1	PTN4_HUMAN	P29074 homo sapien
42	36	50.0	1103	1	VG37_BPAR1	Q9g0b5 bacteriopha
43	36	50.0	1173	1	ATC2_YEAST	P38929 saccharomyc
44	36	50.0	1365	1	YAK1_SCHPO	Q09915 schizosacch
45	36	50.0	3703	1	ABF1_HUMAN	Q15911 homo sapien

ALIGNMENTS

RESULT 1
COBL_METJA
ID COBL_METJA STANDARD; PRT; 211 AA.
AC Q58917;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]
DE (EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y
DE methylase).
GN COBL OR MJ1522.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC ----------

CC ----------
DR EMBL; U67593; AAB99541.1; -.
DR TIGR; MJ1522; -.
DR InterPro; IPR000878; Corrin_porph_mthyltrnf.
DR Pfam; PF00590; TP_methylase; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 211;

```
  Best Local Similarity   54.5%;  Pred. No. 3.8;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 VVGLSPGEQEY 14
           :||: ||::||
Db       4 IVGIGPGDREY 14


RESULT   2
COBI_MYCTU
ID   COBI_MYCTU     STANDARD;      PRT;   508 AA.
AC   Q10677;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE   methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine--precorrin-2
DE   methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN   COBIJ OR RV2066 OR MT2126 OR MTCY49.05.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC       PRECORRIN-3A (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC       adenosyl-L-homocysteine + precorrin-3A.
CC   -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC       AND CBIL/COBI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98214.1; -.
DR   EMBL; AE007063; AAK46406.1; -.
DR   TIGR; MT2126; -.
DR   TubercuList; Rv2066; -.
DR   InterPro; IPR000878; Corrin_porph_mthyltrnf.
DR   InterPro; IPR003043; Uropor_methyltransf.
DR   Pfam; PF00590; TP_methylase; 2.
DR   PROSITE; PS00839; SUMT_1; 1.
DR   PROSITE; PS00840; SUMT_2; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Multifunctional enzyme; Complete proteome.
FT   DOMAIN        1    243       PRECORRIN-2 C20-METHYLTRANSFERASE.
FT   DOMAIN      244    508       PRECORRIN-3 METHYLASE.
SQ   SEQUENCE   508 AA;  53910 MW;  95AC066F022C4DC1 CRC64;


  Query Match             56.9%;  Score 41;  DB 1;  Length 508;
  Best Local Similarity   58.3%;  Pred. No. 20;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 AVVGLSPGEQEY 14
           ||||| ||: ::
Db     250 AVVGLGPGDSDW 261


RESULT   3
KV3I_HUMAN
ID   KV3I_HUMAN     STANDARD;      PRT;   115 AA.
AC   P04433;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-III region VG precursor (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85087932; PubMed=6440122;
RA   Pech M., Zachau H.G.;
RT   "Immunoglobulin genes of different subgroups are interdigitated
RT   within the VK locus.";
RL   Nucleic Acids Res. 12:9229-9236(1984).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01668; -; NOT_ANNOTATED_CDS.
DR   PIR; A01900; K3HUVG.
DR   HSSP; P01789; 1MCP.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-III REGION VG.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109    115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12575 MW;  2DE47CDA3A17D555 CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   63.6%;  Pred. No. 9.9;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQ 12
           || : |||||:
Db      28 PATLSLSPGER 38
```

```
RESULT   4
KV5I_MOUSE
ID   KV5I_MOUSE     STANDARD;      PRT;   115 AA.
AC   P01642;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-V region L7 precursor (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81220975; PubMed=6264318;
RA   Pech M., Hochtl J., Schnell H., Zachau H.G.;
RT   "Differences between germ-line and rearranged immunoglobulin V kappa
RT   coding sequences suggest a localized mutation mechanism.";
RL   Nature 291:668-670(1981).
CC   -!- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPLICE JUNCTIONS AT
CC       THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
CC       LACKING RESIDUES 17-19.
DR   PIR; A01925; KVMSL7.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21   >115       IG KAPPA CHAIN V-V REGION L7.
FT   DOMAIN       21     43       FRAMEWORK-1.
FT   DOMAIN       44     54       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       55     69       FRAMEWORK-2.
FT   DOMAIN       70     76       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       77    108       FRAMEWORK-3.
FT   DOMAIN      109   >115       COMPLEMENTARITY-DETERMINING-3.
FT   DISULFID     43    108       BY SIMILARITY.
FT   NON_TER     115    115
SQ   SEQUENCE   115 AA;  12615 MW;  C17BEC758C577E00 CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 115;
  Best Local Similarity   54.5%;  Pred. No. 9.9;
  Matches    6;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQ 12
           ||:: :||||:
Db      28 PAILSVSPGER 38


RESULT   5
PURU_HAEIN
ID   PURU_HAEIN     STANDARD;      PRT;   278 AA.
AC   Q03432;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE   hydrolase).
GN   PURU OR HI1588.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   SEQUENCE OF 64-278 FROM N.A.
RC   STRAIN=RM 7004;
RX   MEDLINE=93328119; PubMed=8335255;
RA   Maskell D.J.;
RT   "Cloning and sequencing of the Haemophilus influenzae aroA gene.";
RL   Gene 129:155-156(1993).
CC   -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
CC       THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
CC       5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
CC       GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
CC       tetrahydrofolate.
CC   -!- ENZYME REGULATION: ACTIVATED BY METHIONINE, INHIBITED BY GLYCINE
CC       (BY SIMILARITY).
CC   -!- PATHWAY: DE NOVO PURINE BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMOHEXAMER (BY SIMILARITY).
CC   -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U32833; AAC23236.1; -.
DR   EMBL; L04686; AAA24942.1; -.
DR   PIR; PN0606; PN0606.
DR   HSSP; P08179; 1GRC.
DR   TIGR; HI1588; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002376; formyl_transf.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF00551; formyl_transf; 1.
KW   Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW   Complete proteome.
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   CONFLICT    115    117       VIG -> RNR (IN REF. 2).
FT   CONFLICT    138    140       HEN -> PK (IN REF. 2).
FT   CONFLICT    205    205       K -> E (IN REF. 2).
SQ   SEQUENCE   278 AA;  32173 MW;  7F375AB3C422EC4B CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 278;
  Best Local Similarity   50.0%;  Pred. No. 24;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQEY 14
           ||| :|  | :| |
Db     191 LPAFIGAKPYQQAY 204


RESULT   6
C4AD_DROME
ID   C4AD_DROME     STANDARD;      PRT;   516 AA.
AC   Q9V4T3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYPIVAD1).
GN   CYP4AD1 OR CG2110.
```

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC       IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC       oxidized flavoprotein + H(2)O.
CC   -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003837; AAF59092.1; -.
DR   FlyBase; FBgn0033292; Cyp4ad1.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PRINTS; PR00385; P450.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
KW   Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW   Endoplasmic reticulum; Hypothetical protein.
FT   BINDING     445    445       HEME (BY SIMILARITY).
SQ   SEQUENCE   516 AA;  58870 MW;  648EA22492AF58C7 CRC64;
```

```
  Query Match             54.2%;  Score 39;  DB 1;  Length 516;
  Best Local Similarity   57.1%;  Pred. No. 44;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQEY 14
           |||| || ||  ::
Db     469 LPAVDGLPPGINDH 482


RESULT   7
NR41_RAT
ID   NR41_RAT       STANDARD;      PRT;   597 AA.
AC   P22829;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
DE   (NGFI-B) (NUR77).
GN   NR4A1 OR HMR OR NGFIB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90166506; PubMed=3272167;
RA   Milbrandt J.;
RT   "Nerve growth factor induces a gene homologous to the glucocorticoid
RT   receptor gene.";
RL   Neuron 1:183-188(1988).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93361012; PubMed=8395013;
RA   Wilson T.E., Fahrner T.J., Milbrandt J.;
RT   "The orphan receptors NGFI-B and steroidogenic factor 1 establish
RT   monomer binding as a third paradigm of nuclear receptor-DNA
RT   interaction.";
RL   Mol. Cell. Biol. 13:5794-5804(1993).
RN   [3]
RP   DNA BINDING MOTIFS.
RX   MEDLINE=92229411; PubMed=1314418;
RA   Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
RT   "Participation of non-zinc finger residues in DNA binding by two
RT   nuclear orphan receptors.";
RL   Science 256:107-110(1992).
RN   [4]
RP   PHOSPHORYLATION.
RX   MEDLINE=94043340; PubMed=8227042;
RA   Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;
RT   "The phosphorylation and DNA binding of the DNA-binding domain of the
RT   orphan nuclear receptor NGFI-B.";
RL   J. Biol. Chem. 268:24808-24812(1993).
CC   -!- FUNCTION: PROBABLE NUCLEAR RECEPTOR. MAY ACT CONCOMITANTLY WITH
CC       NURR1 IN REGULATING THE EXPRESSION OF DELAYED-EARLY GENES DURING
CC       LIVER REGENERATION. RECOGNIZES THE SINGLE HALF-SITE 5'-AAAAGGTCA-
CC       3'.
CC   -!- SUBUNIT: BINDS DNA AS A MONOMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE LUNG, BRAIN, SUPERIOR
CC       CERVICAL GANGLIA. HIGH LEVELS SEEN IN THE ADRENAL TISSUE.
CC   -!- INDUCTION: BY NERVE GROWTH FACTOR AND DURING LIVER REGENERATION.
CC   -!- PTM: PHOSPHORYLATION OF SER-350 RESULTS IN DECREASE IN NBRE
CC       BINDING WHILE PHOSPHORYLATION OF SER-340 HAS LITTLE EFFECT ON IT.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR4 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17254; AAA56770.1; ALT_INIT.
DR   PIR; JQ0623; JQ0623.
DR   HSSP; P19793; 2NLL.
DR   TRANSFAC; T00619; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Strdhormone_receptor.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Phosphorylation.
FT   DNA_BIND     266     331       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      266     286       C4-TYPE.
FT   ZN_FING      302     326       C4-TYPE.
FT   DOMAIN       408     458       LIGAND-BINDING (POTENTIAL).
FT   MOD_RES      340     340       PHOSPHORYLATION (BY PKA).
FT   MOD_RES      350     350       PHOSPHORYLATION (BY PKA).
FT   MUTAGEN      340     340       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      350     350       S->A: LOSS OF PHOSPHORYLATION.
FT   MUTAGEN      345     345       R->K: DECREASED NBRE BINDING.
FT   MUTAGEN      348     348       L->V: ALMOST COMPLETE LOSS OF NBRE
FT                                  BINDING.
SQ   SEQUENCE   597 AA;  64281 MW;  9CFA987112337E53 CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 597;
  Best Local Similarity   46.2%;  Pred. No. 51;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 LPAVVGLSPGEQE 13
          :|  : ||||:|:
Db    423 IPGFIELSPGDQD 435


RESULT   8
Y197_MOUSE
ID   Y197_MOUSE     STANDARD;      PRT;  1402 AA.
AC   Q9Z0W3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein KIAA0197 (GTL-13).
GN   KIAA0197 OR GTL1-13.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RA   Van de Putte T., Cozijnsen M., Dewulf N., Tylzanowski P., Lonnoy O.,
RA   Huylebroeck D.;
RT   "Mus musculus mRNA for gtl-13 (gene trap locus-13), similar to human
RT   KIAA0197 gene (D83781), complete cds.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF104415; AAD17922.2; -.
DR   MGD; MGI:1926227; Gtl1-13.
SQ   SEQUENCE   1402 AA;  158230 MW;  3BF5D9F057D28772 CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 1402;
  Best Local Similarity   70.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 LPAVVGLSPG 10
          :|:| |||||
Db    170 IPSVPGLSPG 179


RESULT   9
BUD3_YEAST
ID   BUD3_YEAST     STANDARD;      PRT;  1636 AA.
AC   P25558; P25557; P87007;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bud site selection protein BUD3.
GN   BUD3 OR YCL014W OR YCL14W/YCL13W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95247824; PubMed=7730410;
RA   Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT   "Role of Bud3p in producing the axial budding pattern of yeast.";
RL   J. Cell Biol. 129:767-778(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA   Stateva L.I.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVISIONS.
RA   Mewes H.-W.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
CC       COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC       DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC       ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17580; AAA86315.1; -.
DR   EMBL; X59720; CAA42346.2; -.
DR   PIR; S19341; S19341.
DR   PIR; S19340; S19340.
DR   SGD; S0000520; BUD3.
DR   InterPro; IPR000219; RhoGEF.
DR   SMART; SM00325; RhoGEF; 1.
SQ   SEQUENCE   1636 AA;  184717 MW;  9E4E46BA5C3A3F69 CRC64;


  Query Match             54.2%;  Score 39;  DB 1;  Length 1636;
  Best Local Similarity   58.3%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 PAVVGLSPGEQE 13
```

```
          ||:  ||| :||
Db    112 PAIENLSPSDQE 123


RESULT  10
COBL_MYCTU
ID   COBL_MYCTU     STANDARD;      PRT;   390 AA.
AC   Q10671;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Precorrin-6Y C5,15-methyltransferase [decarboxylating] (EC 2.1.1.132)
DE   (Precorrin-6 methyltransferase) (Precorrin-6Y methylase).
GN   COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC       PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC   -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC       S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC   -!- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC   -!- SIMILARITY: TO S.TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
CC       METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z73966; CAA98225.1; -.
DR   EMBL; AE007063; AAK46412.1; -.
DR   TIGR; MT2132; -.
DR   TubercuList; Rv2072c; -.
DR   InterPro; IPR000878; Corrin_porph_mthyltrnf.
DR   InterPro; IPR000051; SAM_bind.
DR   Pfam; PF00590; TP_methylase; 1.
KW   Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW   Methyltransferase; Complete proteome.
FT   CONFLICT    205    205       L -> P (IN REF. 2).
FT   CONFLICT    327    327       D -> H (IN REF. 2).
SQ   SEQUENCE   390 AA;  41854 MW;  FB42EFF7562F21F3 CRC64;

  Query Match           52.8%;  Score 38;  DB 1;  Length 390;
  Best Local Similarity   88.9%;  Pred. No. 49;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSP 9
           |||| ||||
Db      55 LPAVQGLSP 63


RESULT  11
RS1_LEULA
ID   RS1_LEULA      STANDARD;      PRT;   429 AA.
AC   P50889; P71450;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   40S ribosomal protein S1.
GN   RPS1.
OS   Leuconostoc lactis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Leuconostoc.
OX   NCBI_TaxID=1246;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97186703; PubMed=9034319;
RA   Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT   "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT   shares common epitopes and common DNA binding properties with a
RT   mammalian DNA binding nuclear factor.";
RL   Gene 185:99-103(1997).
RN   [2]
RP   SEQUENCE OF 24-429 FROM N.A.
RX   MEDLINE=95237615; PubMed=7721096;
RA   Eklund E.A., Lee S.W., Skalnik D.G.;
RT   "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT   ribosomal protein S1.";
RL   Gene 155:231-235(1995).
RN   [3]
RP   SEQUENCE OF 78-429 FROM N.A.
RX   MEDLINE=96164600; PubMed=8568274;
RA   Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA   Winkler T.H., Kalden J.R., Reichlin M.;
RT   "Lupus autoantibodies to double-stranded DNA cross-react with
RT   ribosomal protein S1.";
RL   J. Immunol. 156:1668-1675(1996).
CC   -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC       DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC   -!- SIMILARITY: BELONGS TO THE S1P FAMILY OF RIBOSOMAL PROTEINS.
CC   -!- SIMILARITY: CONTAINS 4 S1 MOTIF DOMAINS.
CC   -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC       FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC       CONTAMINATION BY L.LACTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U24086; AAB08978.1; -.
DR   EMBL; U05589; AAA77669.1; -.
DR   EMBL; U27517; AAA97575.1; -.
DR   HSSP; P05055; 1SRO.
DR   InterPro; IPR000110; Ribosomal_S1.
DR   InterPro; IPR003029; S1.
DR   Pfam; PF00575; S1; 4.
DR   PRINTS; PR00681; RIBOSOMALS1.
DR   SMART; SM00316; S1; 4.
DR   PROSITE; PS50126; S1; 4.
KW   Ribosomal protein; Repeat; RNA-binding.
```

```
FT   DOMAIN       55    128       S1 MOTIF 1.
FT   DOMAIN      144    211       S1 MOTIF 2.
FT   DOMAIN      231    299       S1 MOTIF 3.
FT   DOMAIN      316    385       S1 MOTIF 4.
FT   CONFLICT     24     24       S -> G (IN REF. 2).
FT   CONFLICT    122    122       A -> S (IN REF. 3).
FT   CONFLICT    217    217       L -> R (IN REF. 2 AND 3).
SQ   SEQUENCE   429 AA;  46386 MW;  92AC82605F39DDFC CRC64;


  Query Match             52.8%;  Score 38;  DB 1;  Length 429;
  Best Local Similarity   80.0%;  Pred. No. 54;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 AVVGLSPGEQ 12
           |||||| ||:
Db      71 AVVGLSTGEE 80


RESULT  12
CATA_PICAN
ID   CATA_PICAN     STANDARD;      PRT;   507 AA.
AC   P30263;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Catalase (EC 1.11.1.6).
GN   PXP9 OR PXP-9.
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 34438;
RX   MEDLINE=92299073; PubMed=1607006;
RA   Didion T., Roggenkamp R.O.;
RT   "Targeting signal of the peroxisomal catalase in the methylotrophic
RT   yeast Hansenula polymorpha.";
RL   FEBS Lett. 303:113-116(1992).
CC   -!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
CC       SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
CC       PEROXIDE.
CC   -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC   -!- COFACTOR: HEME GROUP.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: Peroxisomal.
CC   -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56501; CAA39856.1; -.
DR   PIR; S23422; S23422.
DR   HSSP; P15202; 1A4E.
DR   InterPro; IPR002226; Catalase.
DR   Pfam; PF00199; catalase; 1.
DR   PRINTS; PR00067; CATALASE.
DR   ProDom; PD000510; Catalase; 1.
DR   PROSITE; PS00342; MICROBODIES_CTER; 1.
DR   PROSITE; PS00437; CATALASE_1; 1.
DR   PROSITE; PS00438; CATALASE_2; 1.
KW   Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW   Peroxisome.
FT   ACT_SITE     65     65       BY SIMILARITY.
FT   ACT_SITE    138    138       BY SIMILARITY.
FT   BINDING     348    348       PROXIMAL HEME LIGAND (BY SIMILARITY).
FT   SITE        505    507       MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ   SEQUENCE   507 AA;  57849 MW;  3536ED0A49539CC3 CRC64;


  Query Match             52.8%;  Score 38;  DB 1;  Length 507;
  Best Local Similarity   70.0%;  Pred. No. 64;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 VVGLSPGEQE 13
           |:| :|||||
Db     445 VLGRTPGEQE 454


RESULT  13
EFG_HAEIN
ID   EFG_HAEIN      STANDARD;      PRT;   699 AA.
AC   P43925;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Elongation factor G (EF-G).
GN   FUSA OR FUS OR HI0579.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC       OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC       RIBOSOME.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U32739; AAC22237.1; -.
DR   HSSP; P13551; 1FNM.
DR   TIGR; HI0579; -.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR000795; GTP_EFTU.
DR   InterPro; IPR004161; GTP_EFTU_D2.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   NP_BIND      16     23       GTP (BY SIMILARITY).
FT   NP_BIND      87     91       GTP (BY SIMILARITY).
FT   NP_BIND     141    144       GTP (BY SIMILARITY).
```

```
SQ   SEQUENCE   699 AA;  77132 MW;  FBBADE39C0F62801 CRC64;


  Query Match             52.8%;  Score 38;  DB 1;  Length 699;
  Best Local Similarity   46.2%;  Pred. No. 88;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQE 13
           :||: |::| | |
Db     291 IPAIKGINPDETE 303


RESULT  14
EFG_PASMU
ID   EFG_PASMU      STANDARD;      PRT;   700 AA.
AC   P57938;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Elongation factor G (EF-G).
GN   FUSA OR PM1356.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION OF
CC       THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC       RIBOSOME.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006173; AAK03440.1; -.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR000795; GTP_EFTU.
DR   InterPro; IPR004161; GTP_EFTU_D2.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   NP_BIND      16     23       GTP (BY SIMILARITY).
FT   NP_BIND      87     91       GTP (BY SIMILARITY).
FT   NP_BIND     141    144       GTP (BY SIMILARITY).
SQ   SEQUENCE   700 AA;  77186 MW;  6CC161F7F9FA9C72 CRC64;


  Query Match             52.8%;  Score 38;  DB 1;  Length 700;
  Best Local Similarity   46.2%;  Pred. No. 88;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQE 13
           :||: |::| | |
Db     292 IPAIKGINPDETE 304


RESULT  15
PMXA_MOUSE
ID   PMXA_MOUSE     STANDARD;      PRT;   280 AA.
AC   Q62066;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Paired mesoderm homeobox protein 2A (Paired-like homeobox 2A)
DE   (PHOX2A homeodomain protein) (Aristaless homeobox protein homolog).
GN   ARIX OR PMX2A OR PMX2 OR PHOX2A OR PHOX2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/J;
RX   MEDLINE=94244481; PubMed=7910552;
RA   Valarche T.S., Tissier-Seta J.P., Hirsch M.R., Martinez S.F.,
RA   Goridis C., Brunet J.F.;
RT   "The mouse homeodomain protein Phox2 regulates Ncam promoter activity
RT   in concert with Cux/CDP and is a putative determinant of
RT   neurotransmitter phenotype.";
RL   Development 119:881-896(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN REGULATING THE SPECIFICITY OF
CC       EXPRESSION OF THE CATECHOLAMINE BIOSYNTHETIC GENES. ACTS AS A
CC       TRANSCRIPTION ACTIVATOR/FACTOR. COULD MAINTAIN THE NORADRENERGIC
CC       PHENOTYPE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X75014; CAA52923.1; -.
DR   HSSP; P22808; 1NK3.
DR   MGD; MGI:106633; Arix.
DR   InterPro; IPR000047; HTH_repressr.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     90    149       HOMEOBOX.
FT   DOMAIN      193    198       POLY-PRO.
FT   DOMAIN      233    242       POLY-GLY.
SQ   SEQUENCE   280 AA;  29417 MW;  F7CBFBCE66B57A5D CRC64;


  Query Match             51.4%;  Score 37;  DB 1;  Length 280;
  Best Local Similarity   58.3%;  Pred. No. 52;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQ 12
           ||| :|  || |
Db     210 LPAALGSGPGPQ 221


Search completed: September  9, 2002, 06:56:00
Job time: 333 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:49:12 ; Search time 77.59 Seconds
(without alignments)
31.214 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	44	61.1	381	16	Q9RRL5	Q9rrl5 deinococcus
2	43	59.7	633	3	Q9Y7N6	Q9y7n6 schizosacch
3	43	59.7	821	17	Q9HPR8	Q9hpr8 halobacteri
4	41	56.9	156	16	Q9A5K8	Q9a5k8 caulobacter
5	41	56.9	228	17	Q9HKE3	Q9hke3 thermoplasm
6	41	56.9	326	1	Q9UXP0	Q9uxp0 methanolobu
7	41	56.9	11096	2	Q9L4W3	Q9l4w3 streptomyce
8	40	55.6	227	17	Q97WD1	Q97wd1 sulfolobus
9	40	55.6	355	10	Q947A7	Q947a7 nitellopsis
10	40	55.6	358	4	Q9NWD0	Q9nwd0 homo sapien
11	40	55.6	386	4	Q9NWM3	Q9nwm3 homo sapien
12	40	55.6	428	16	Q92AG9	Q92ag9 listeria in
13	40	55.6	540	16	Q9RR71	Q9rr71 deinococcus
14	40	55.6	661	10	Q9SKD1	Q9skd1 arabidopsis
15	40	55.6	673	3	Q9P327	Q9p327 schizosacch
16	40	55.6	683	2	O34302	O34302 rhizobium e
17	39	54.2	277	16	Q9KQK6	Q9kqk6 vibrio cho
18	39	54.2	407	2	Q9LCW0	Q9lcw0 streptomyce
19	38.5	53.5	656	16	Q9RW14	Q9rw14 deinococcus
20	38	52.8	136	11	Q61061	Q61061 mus musculu
21	38	52.8	152	17	Q9V2S2	Q9v2s2 pyrococcus
22	38	52.8	155	17	O57778	O57778 pyrococcus
23	38	52.8	200	16	Q97LH7	Q97lh7 clostridium
24	38	52.8	212	5	Q9NA60	Q9na60 caenorhabdi
25	38	52.8	222	17	Q97WC8	Q97wc8 sulfolobus
26	38	52.8	280	12	O12662	O12662 broad bean
27	38	52.8	325	16	Q9CBS2	Q9cbs2 mycobacteri
28	38	52.8	409	5	Q9N5W7	Q9n5w7 caenorhabdi
29	38	52.8	519	10	O48655	O48655 oryza sativ
30	38	52.8	540	12	Q9IV49	Q9iv49 human calic
31	38	52.8	540	12	Q918B0	Q918b0 norwalk-lik
32	38	52.8	663	9	Q94MX4	Q94mx4 bacteriopha
33	38	52.8	1035	12	Q91TP4	Q91tp4 tupaia herp
34	38	52.8	1317	11	O35211	O35211 mus musculu
35	38	52.8	1769	12	Q9WP48	Q9wp48 broad bean
36	38	52.8	1804	11	Q9Z1Q1	Q9z1q1 mesocricetu
37	38	52.8	1870	12	Q9YM10	Q9ym10 broad bean
38	38	52.8	2152	2	Q9ALM5	Q9alm5 saccharopol
39	38	52.8	4823	13	O93321	O93321 fugu rubrip
40	38	52.8	5588	2	Q9ALM2	Q9alm2 saccharopol
41	38	52.8	6146	2	Q93HJ5	Q93hj5 streptomyce
42	38	52.8	8817	2	Q53840	Q53840 polyangium
43	37.5	52.1	451	10	Q9XHE0	Q9xhe0 zea mays (m
44	37.5	52.1	827	11	P70218	P70218 mus musculu
45	37.5	52.1	1804	1	Q9P9A9	Q9p9a9 uncultured

ALIGNMENTS

RESULT 1
Q9RRL5
ID Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CYTOCHROME P450.
GN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AE002076; AAF12016.1; -.
DR TIGR; DR2473; -.
DR InterPro; IPR001128; Cyt_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 61.1%; Score 44; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 19;

```
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LPAVVGLSP 9
          |||||||||
Db     52 LPAVVGLSP 60


RESULT   2
Q9Y7N6
ID   Q9Y7N6      PRELIMINARY;      PRT;   633 AA.
AC   Q9Y7N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LYSOPHOSPHOLIPASE C1450.09C PRECURSOR (EC 3.1.1.5).
GN   SPCC1450.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Volckaert G.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES THE RELEASE OF FATTY ACIDS FROM
CC       LYSOPHOSPHOLIPIDS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 2-LYSOPHOSPHATIDYLCHOLINE + H(2)O =
CC       GLYCEROPHOSPHOCHOLINE + A FATTY ACID ANION.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: TO OTHER FUNGAL LYSOPHOSPHOLIPASES AND TO YEAST
CC       SPO1.
DR   EMBL; AL049559; CAB40176.2; -.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
KW   Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    633       LYSOPHOSPHOLIPASE C1450.09C.
FT   DOMAIN      404    407       POLY-THR.
FT   CARBOHYD    118    118       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    264    264       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    360    360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    367    367       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    508    508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    537    537       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    564    564       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   633 AA;  68292 MW;  49871B2955893D19 CRC64;


  Query Match           59.7%;  Score 43;  DB 3;  Length 633;
  Best Local Similarity   69.2%;  Pred. No. 49;
  Matches    9;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 PAVVGLSPGEQEY 14
          ||  ||| ||||:
Db     71 PASDGLSTGEQEF 88


RESULT   3
Q9HPR8
ID   Q9HPR8      PRELIMINARY;      PRT;   821 AA.
AC   Q9HPR8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DNA HELICASE.
GN   HEL OR VNG1501G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005064; AAG19799.1; -.
DR   InterPro; IPR001478; PDZ.
DR   SMART; SM00228; PDZ; 1.
KW   Helicase; Complete proteome.
SQ   SEQUENCE   821 AA;  89848 MW;  C454C76B984A5702 CRC64;


  Query Match           59.7%;  Score 43;  DB 17;  Length 821;
  Best Local Similarity   90.0%;  Pred. No. 66;
  Matches    9;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 AVVGLSPGEQ 12
          ||||||| ||
Db    326 AVVGLSPAEQ 335


RESULT   4
Q9A5K8
ID   Q9A5K8      PRELIMINARY;      PRT;   156 AA.
AC   Q9A5K8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN CC2439.
GN   CC2439.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005913; AAK24410.1; -.
DR   TIGR; CC2439; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   156 AA;  17142 MW;  21F54D8648396141 CRC64;
```

```
  Query Match             56.9%;  Score 41;  DB 16;  Length 156;
  Best Local Similarity   80.0%;  Pred. No. 24;
  Matches    8;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 AVVGLSPGEQ 12
          ||||| |||:
Db     18 AVVGLDPGEK 27


RESULT   5
Q9HKE3
ID   Q9HKE3      PRELIMINARY;      PRT;   228 AA.
AC   Q9HKE3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PRECORRIN-2 METHYLTRANSFERASE RELATED PROTEIN.
GN   TA0658.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445065; CAC11796.1; -.
DR   InterPro; IPR000878; Corrin_porph_mthyltrnf.
DR   InterPro; IPR003043; Uropor_methyltransf.
DR   Pfam; PF00590; TP_methylase; 1.
DR   PROSITE; PS00839; SUMT_1; UNKNOWN_1.
KW   Transferase; Methyltransferase; Complete proteome.
SQ   SEQUENCE   228 AA;  25084 MW;  11ABD8B68192A67C CRC64;


  Query Match             56.9%;  Score 41;  DB 17;  Length 228;
  Best Local Similarity   63.6%;  Pred. No. 36;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 VVGLSPGEQEY 14
          |||| ||: :|
Db      5 VVGLGPGDPDY 15


RESULT   6
Q9UXP0
ID   Q9UXP0      PRELIMINARY;      PRT;   326 AA.
AC   Q9UXP0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F420-DEPENDENT N5,N10-METHYLENE-TETRAHYDROMETHANOPTERIN REDUCTASE,
DE   PUTATIVE.
GN   FFDA.
OS   Methanolobus tindarius.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanolobus.
OX   NCBI_TaxID=2221;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 2278;
RA   Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA   Gottschalk G., Blaut M.;
RT   "The F420H2-dehydrogenase from Methanolobus tindarius: Cloning of the
RT   ffd operon and expression of the genes in Escherichia coli.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ011519; CAB56639.1; -.
DR   InterPro; IPR002103; Bac_luciferase.
DR   Pfam; PF00296; bac_luciferase; 1.
SQ   SEQUENCE   326 AA;  34043 MW;  16F3AB9733A45D82 CRC64;


  Query Match             56.9%;  Score 41;  DB 1;  Length 326;
  Best Local Similarity   70.0%;  Pred. No. 53;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 AVVGLSPGEQ 12
          |::|| ||||
Db     88 AILGLGPGEQ 97


RESULT   7
Q9L4W3
ID   Q9L4W3      PRELIMINARY;      PRT; 11096 AA.
AC   Q9L4W3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NYSC.
GN   NYSC.
OS   Streptomyces noursei.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1971;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 11455;
RX   MEDLINE=20334850; PubMed=10873841;
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT   Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT   deduction of the biosynthetic pathway.";
RL   Chem. Biol. 7:395-403(2000).
DR   EMBL; AF263912; AAF71776.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR001227; Acyltransf_domain.
DR   InterPro; IPR002085; Adh_zn_family.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Phosphopant_attach.
DR   Pfam; PF00698; Acyl_transf; 6.
DR   Pfam; PF00107; adh_zinc; 1.
DR   Pfam; PF00109; ketoacyl-synt; 6.
DR   Pfam; PF02801; ketoacyl-synt_C; 6.
DR   PROSITE; PS50075; ACP_DOMAIN; 6.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW   Phosphopantetheine; Transferase.
SQ   SEQUENCE   11096 AA;  1150415 MW;  776CAEAFCAE551DD CRC64;


  Query Match             56.9%;  Score 41;  DB 2;  Length 11096;
  Best Local Similarity   63.6%;  Pred. No. 2.5e+03;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 PAVVGLSPGEQ 12
          | | ||:||:|
Db   4998 PEVTGLAPGDQ 5008


RESULT   8
Q97WD1
ID   Q97WD1      PRELIMINARY;      PRT;   227 AA.
AC   Q97WD1;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   COBALAMIN BIOSYNTHESIS PRECORRIN-6B METHYLASE, PUTATIVE
DE   (CBIE).
GN   CBIE.
```

```
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR   EMBL; AE006833; AAK42457.1; -.
DR   InterPro; IPR000878; Corrin_porph_mthyltrnf.
DR   Pfam; PF00590; TP_methylase; 1.
KW   Methyltransferase; Complete proteome.
SQ   SEQUENCE   227 AA;  25548 MW;  0267F09491F2DCED CRC64;


  Query Match             55.6%;  Score 40;  DB 17;  Length 227;
  Best Local Similarity   54.5%;  Pred. No. 53;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 VVGLSPGEQEY 14
           :||: ||: ||
Db      10 IVGVGPGDPEY 20


RESULT   9
Q947A7
ID   Q947A7      PRELIMINARY;      PRT;   355 AA.
AC   Q947A7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENOLASE (FRAGMENT).
GN   ENO.
OS   Nitellopsis obtusa.
OC   Eukaryota; Viridiplantae; Streptophyta; Characeae; Nitellopsis.
OX   NCBI_TaxID=40811;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X-854;
RX   MEDLINE=21437986; PubMed=11526220;
RA   Keeling P.J., Palmer J.D.;
RT   "Lateral transfer at the gene and subgenic levels in the evolution of
RT   eukaryotic enolase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
DR   EMBL; AF348916; AAL05455.1; -.
FT   NON_TER       1      1
FT   NON_TER     355    355
SQ   SEQUENCE   355 AA;  38295 MW;  400DF160087DE450 CRC64;


  Query Match             55.6%;  Score 40;  DB 10;  Length 355;
  Best Local Similarity   54.5%;  Pred. No. 87;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQ 12
           |||:|: | :|
Db      28 PAVIGMDPADQ 38


RESULT   10
Q9NWD0
ID   Q9NWD0      PRELIMINARY;      PRT;   358 AA.
AC   Q9NWD0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 38.8 KDA PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000977; BAA91452.1; -.
DR   InterPro; IPR003892; CUE.
DR   Pfam; PF02845; CUE; 1.
DR   SMART; SM00546; CUE; 1.
SQ   SEQUENCE   358 AA;  38760 MW;  BCE1AA7E95C73BF0 CRC64;


  Query Match             55.6%;  Score 40;  DB 4;  Length 358;
  Best Local Similarity   58.3%;  Pred. No. 87;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQE 13
           ||: |  ||:||
Db     213 PAMAGPGPGDQE 224


RESULT   11
Q9NWM3
ID   Q9NWM3      PRELIMINARY;      PRT;   386 AA.
AC   Q9NWM3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ20739 FIS, CLONE HEP07341 (HYPOTHETICAL 42.3 KDA
DE   PROTEIN).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA   Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000746; BAA91357.1; -.
DR   EMBL; BC008034; AAH08034.1; -.
DR   InterPro; IPR003892; CUE.
DR   Pfam; PF02845; CUE; 1.
DR   SMART; SM00546; CUE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   386 AA;  42258 MW;  F5530FA47C267895 CRC64;


  Query Match             55.6%;  Score 40;  DB 4;  Length 386;
  Best Local Similarity   58.3%;  Pred. No. 95;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQE 13
```

```
          ||: |  ||:||
Db    213 PAMAGPGPGDQE 224


RESULT  12
Q92AG9
ID   Q92AG9      PRELIMINARY;      PRT;   428 AA.
AC   Q92AG9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PYRP PROTEIN.
GN   PYRP OR LIN1953.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596170; CAC97183.1; -.
DR   ListiList; LIN01953; -.
KW   Complete proteome.
SQ   SEQUENCE   428 AA;  44835 MW;  4A2D93B1FA95CCE7 CRC64;


  Query Match             55.6%;  Score 40;  DB 16;  Length 428;
  Best Local Similarity   70.0%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPG 10
           :|:| |||||
Db      42 VPSVTGLSPG 51


RESULT  13
Q9RR71
ID   Q9RR71      PRELIMINARY;      PRT;   540 AA.
AC   Q9RR71;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CARBOXYLESTERASE, TYPE B.
GN   DR2626.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
DR   EMBL; AE002092; AAF12163.1; -.
DR   HSSP; P37967; 1QE3.
DR   MEROPS; S09.UNC; -.
DR   TIGR; DR2626; -.
DR   InterPro; IPR002018; Carboxylesterase_B.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00135; COesterase; 1.
DR   PROSITE; PS00122; CARBOXYLESTERASE_B_1; 1.
KW   Complete proteome; Hydrolase.
SQ   SEQUENCE   540 AA;  55899 MW;  CDB97C69E288EFD5 CRC64;


  Query Match             55.6%;  Score 40;  DB 16;  Length 540;
  Best Local Similarity   70.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGE 11
           | |:||:|||
Db     512 PQVLGLAPGE 521


RESULT  14
Q9SKD1
ID   Q9SKD1      PRELIMINARY;      PRT;   661 AA.
AC   Q9SKD1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G13900 PROTEIN.
GN   AT2G13900.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006528; AAF18650.1; -.
DR   InterPro; IPR004146; DC1.
DR   InterPro; IPR001965; PHD.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF03107; DC1; 2.
DR   SMART; SM00249; PHD; 4.
DR   SMART; SM00184; RING; 2.
SQ   SEQUENCE   661 AA;  76822 MW;  5CE1FD9F848A60D7 CRC64;


  Query Match             55.6%;  Score 40;  DB 10;  Length 661;
  Best Local Similarity   53.8%;  Pred. No. 1.7e+02;
  Matches    7;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQEY 14
           |  : : ||||||
```

```
Db    437 PLTLKIVPGEQEY 449


RESULT  15
Q9P327
ID   Q9P327      PRELIMINARY;      PRT;   673 AA.
AC   Q9P327;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE LYSOPHOSPHOLIPASE PRECURSOR.
GN   SPAC1348.10C OR SPAC977.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Lyne M., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Zimmermann W., Wambutt R., McDougall R.C., Rajandream M.A.,
RA   Barrell B.G.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL358912; CAB94277.1; -.
DR   EMBL; AL137130; CAB69631.1; -.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   InterPro; IPR002642; PLAc.
DR   Pfam; PF01735; PLA2_B; 1.
DR   SMART; SM00022; PLAc; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   673 AA;  74595 MW;  B39A773E76CD694B CRC64;


  Query Match             55.6%;  Score 40;  DB 3;  Length 673;
  Best Local Similarity   61.5%;  Pred. No. 1.7e+02;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 PAVVGLSPGEQEY 14
           ||  ||: ||| |
Db      83 PASEGLNEGEQSY 95


Search completed: September  9, 2002, 06:55:26
Job time: 374 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 9, 2002, 06:36:12 ; Search time 64.82 Seconds
(without alignments)
23.990 Million cell updates/sec

Title: US-09-786-214-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGEQEY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	100.0	14	21	AAY84266	Peptide derived fr
2	72	100.0	20	21	AAY84265	Truncated macropha
3	72	100.0	25	21	AAY84264	Peptide of alterna
4	68	94.4	13	21	AAY84267	Peptide derived fr
5	65	90.3	13	21	AAY84268	Peptide derived fr
6	65	90.3	15	21	AAY84269	Peptide derived fr
7	45	62.5	234	22	AAB36208	Human immune syste
8	41	56.9	306	22	ABG19552	Novel human diagno
9	41	56.9	11096	22	AAE10129	Streptomyces nours
10	40	55.6	261	22	AAG91441	C glutamicum prote
11	40	55.6	306	22	ABG24698	Novel human diagno
12	40	55.6	358	22	AAB92530	Human protein sequ
13	40	55.6	668	21	AAB57055	Human prostate can
14	39	54.2	18	20	AAY41875	Rheumatoid arthrit
15	39	54.2	18	22	AAU25388	Schizophrenia-Asso
16	39	54.2	18	22	AAU15732	Schizophrenia-asso
17	39	54.2	88	21	AAY56655	Partial peptide fr
18	39	54.2	104	18	AAW31723	Alpha light chain
19	39	54.2	104	19	AAW26795	Anti-gp54 MAb 48-1
20	39	54.2	106	14	AAR41234	Monoclonal antibod
21	39	54.2	106	19	AAW71241	Light chain variab
22	39	54.2	107	14	AAR32129	Anti-IL2R beta ant
23	39	54.2	107	14	AAR37610	B-B10 MAb L chain
24	39	54.2	107	15	AAR50190	Light chain variab
25	39	54.2	107	18	AAW28531	Humanised cA2 ligh
26	39	54.2	107	18	AAW24513	Anti-Factor-IX hum
27	39	54.2	107	18	AAW24514	Anti-Factor-IX hum
28	39	54.2	107	18	AAW24515	Anti-Factor-IX hum
29	39	54.2	107	18	AAW08948	Kappa light chain
30	39	54.2	107	18	AAW08949	Kappa light chain
31	39	54.2	107	19	AAW76006	LM609 grafted anti
32	39	54.2	107	19	AAW76002	Vitaxin antibody l
33	39	54.2	107	19	AAW40820	Light chain variab
34	39	54.2	107	20	AAY26979	Light chain variab
35	39	54.2	107	20	AAY30203	The variable light
36	39	54.2	107	20	AAY30205	The variable light
37	39	54.2	107	20	AAY23243	Light chain variab
38	39	54.2	107	21	AAB10023	H. pylori beta-ure
39	39	54.2	107	22	AAG79125	Amino acid sequenc
40	39	54.2	107	22	AAE10849	Mouse light chain
41	39	54.2	107	22	AAG67762	Anti-TNF antibody
42	39	54.2	107	22	AAG63588	A light chain vari
43	39	54.2	107	22	AAB85350	Antibody AAL 160 l
44	39	54.2	107	22	AAB86071	H. pylori beta-ure
45	39	54.2	107	22	AAB86109	H. pylori beta-ure

ALIGNMENTS

RESULT 1
AAY84266
ID AAY84266 standard; Peptide; 14 AA.
XX
AC AAY84266;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue

PT sequence -
XX
PS Claim 2; Page 39; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 14 AA;

Query Match 100.0%; Score 72; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQEY 14
||||||||||||||
Db 1 lpavvglspgeqey 14

RESULT 2
AAY84265
ID AAY84265 standard; Peptide; 20 AA.
XX
AC AAY84265;
XX
DT 12-JUL-2000 (first entry)
XX
DE Truncated macrophage colony stimulating factor tumour antigen.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99675.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Claim 3; Page 64; 74pp; English.
XX
CC The present sequence represents a truncated tumour rejection antigen
CC precursor, and is encoded by a truncated alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 20 AA;

Query Match 100.0%; Score 72; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQEY 14
||||||||||||||
Db 4 lpavvglspgeqey 17

RESULT 3
AAY84264
ID AAY84264 standard; Peptide; 25 AA.
XX
AC AAY84264;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide of alternate reading frame of macrophage colony stimulating gene
XX
KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
DR N-PSDB; AAZ99672.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Claim 1; Page 64; 74pp; English.
XX
CC The present sequence represents a tumour rejection antigen precursor,
CC and is encoded by an alternative open reading frame (ORF) of human
CC macrophage colony stimulating gene. Peptides derived from the
CC alternative ORF of macrophage-colony stimulating factor, when presented
CC by an antigen presenting cell having a human leukocyte antigen (HLA)
CC class I molecule, effectively induce the activation and proliferation
CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids
CC derived from the alternate ORF of macrophage-colony stimulating factor
CC are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 25 AA;

```
  Query Match           100.0%;  Score 72;  DB 21;  Length 25;
  Best Local Similarity  100.0%;  Pred. No. 5.7e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LPAVVGLSPGEQEY 14
          ||||||||||||||
Db      4 lpavvglspgeqey 17


RESULT   4
AAY84267
ID   AAY84267 standard; Peptide; 13 AA.
XX
AC   AAY84267;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US20344.
XX
PR   04-SEP-1998;   98US-0099077.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame
PT   macrophage-colony stimulating factor comprises 25 amino acid residue
PT   sequence -
XX
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics.
XX
SQ   Sequence   13 AA;


  Query Match           94.4%;  Score 68;  DB 21;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 0.00012;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 PAVVGLSPGEQEY 14
          |||||||||||||
Db      1 pavvglspgeqey 13


RESULT   5
AAY84268
ID   AAY84268 standard; Peptide; 13 AA.
XX
AC   AAY84268;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW   tumour rejection antigen; macrophage colony stimulating gene;
KW   macrophage-colony stimulating factor; antigen presenting cell;
KW   human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200013699-A1.
XX
PD   16-MAR-2000.
XX
PF   03-SEP-1999;   99WO-US20344.
XX
PR   04-SEP-1998;   98US-0099077.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Probst-Kepper M,  Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-256859/22.
XX
PT   Isolated polypeptide used to treat subjects having a disorder
PT   characterized by expression of alternative open reading frame
PT   macrophage-colony stimulating factor comprises 25 amino acid residue
PT   sequence -
XX
PS   Example 2; Page 40; 74pp; English.
XX
CC   The present sequence represents a peptide which is derived from a tumour
CC   rejection antigen precursor encoded by an alternative open reading frame
CC   (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC   the alternative ORF of macrophage-colony stimulating factor, when
CC   presented by an antigen presenting cell having a human leukocyte antigen
CC   (HLA) class I molecule, effectively induce the activation and
CC   proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC   acids derived from the alternate ORF of macrophage-colony stimulating
CC   factor are useful for enriching selectively a population of T lymphocytes
CC   with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC   characterized by expression of the polypeptide, and for identifying
CC   functional variants and mimetics.
XX
SQ   Sequence   13 AA;


  Query Match           90.3%;  Score 65;  DB 21;  Length 13;
  Best Local Similarity  100.0%;  Pred. No. 0.00039;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 LPAVVGLSPGEQE 13
          |||||||||||||
Db      1 lpavvglspgeqe 13


RESULT   6
AAY84269
ID   AAY84269 standard; Peptide; 15 AA.
XX
AC   AAY84269;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
```

KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US20344.
XX
PR 04-SEP-1998; 98US-0099077.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame
PT macrophage-colony stimulating factor comprises 25 amino acid residue
PT sequence -
XX
PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics.
XX
SQ Sequence 15 AA;

Query Match 90.3%; Score 65; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
||||||||||||| |
Db 3 lpavvglspgeqe 15

RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
XX
AC AAB36208;
XX
DT 15-FEB-2001 (first entry)
XX
DE Human immune system associated protein HISAP-6.
XX
KW Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.
XX
OS Homo sapiens.
XX
PN US6135941-A.
XX
PD 24-OCT-2000.
XX
PF 27-MAR-1998; 98US-0049672.
XX
PR 27-MAR-1998; 98US-0049672.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
PI Hillman JL, Au-Young J;
XX
DR WPI; 2001-030926/04.
DR N-PSDB; AAC66524.
XX
PT New human immune system associated proteins (HISAP) and polynucleotides
PT encoding the HISAP, useful for diagnosing, treating or preventing
PT immune or cell proliferative disorders or infections -
XX
PS Claim 1; Column 59-60; 54pp; English.
XX
CC The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer.
XX
SQ Sequence 234 AA;

Query Match 62.5%; Score 45; DB 22; Length 234;
Best Local Similarity 72.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
||:| |||||:
Db 28 paivslspger 38

RESULT 8
ABG19552
ID ABG19552 standard; Protein; 306 AA.
XX
AC ABG19552;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19543.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS83739.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

```
XX
PS   Claim 20; SEQ ID No 49911; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   306 AA;


  Query Match             56.9%;  Score 41;  DB 22;  Length 306;
  Best Local Similarity   57.1%;  Pred. No. 1e+02;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 LPAVVGLSPGEQEY 14
           ||   ||||| | :
Db     137 lpqapglspgaqsw 150


RESULT   9
AAE10129
ID   AAE10129 standard; Protein; 11096 AA.
XX
AC   AAE10129;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin gene, NysC.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; PKS type I.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   Domain          35..455
FT                   /label= KS3_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          546..858
FT                   /label= AT3_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          872..1073
FT                   /label= DH3_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          1381..1628
FT                   /label= KR3_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          1662..1735
FT                   /label= ACP3_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          1757..2180
FT                   /label= KS4_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          2291..2603
FT                   /label= AT4_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          2617..2818
FT                   /label= DH4_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          3124..3371
FT                   /label= KR4_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          3407..3480
FT                   /label= ACP4_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          3501..3924
FT                   /label= KS5_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          4032..4346
FT                   /label= AT5_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          4360..4561
FT                   /label= DH5_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          4953..5239
FT                   /label= ER5_domain
FT                   /note= "Enoylreductase (ER) domain"
FT   Domain          5248..5495
FT                   /label= KR5_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          5528..5601
FT                   /label= ACP5_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          5623..6046
FT                   /label= KS6_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          6166..6478
FT                   /label= AT6_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          6492..6704
FT                   /label= DH6_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          7038..7281
FT                   /label= KR6_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          7315..7388
FT                   /label= ACP6_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          7408..7831
FT                   /label= KS7_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          7939..8253
FT                   /label= AT7_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          8267..8470
FT                   /label= DH7_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          8812..9086
FT                   /label= KR7_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          9120..9193
FT                   /label= ACP7_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
FT   Domain          9214..9637
FT                   /label= KS8_domain
FT                   /note= "Ketosynthase (KS) domain"
FT   Domain          9758..10072
FT                   /label= AT8_domain
FT                   /note= "Acyltransferase (AT) domain"
FT   Domain          10086..10289
FT                   /label= DH8_domain
FT                   /note= "Dehydratase (DH) domain"
FT   Domain          10657..10904
FT                   /label= KR8_domain
FT                   /note= "Ketoreductase (KR) domain"
FT   Domain          10939..11012
FT                   /label= ACP8_domain
FT                   /note= "Acyl carrier protein (ACP) domain"
```

XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR N-PSDB; AAD17184.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 15; Page 170-176; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS type I encoding Streptomyces noursei nystatin
CC gene, NysC.
XX
SQ Sequence 11096 AA;

Query Match 56.9%; Score 41; DB 22; Length 11096;
Best Local Similarity 63.6%; Pred. No. 5.4e+03;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
| | ||:||:|
Db 4998 pevtglapgdq 5008

RESULT 10
AAG91441
ID AAG91441 standard; Protein; 261 AA.
XX
AC AAG91441;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5195.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH66660.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 5195; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 261 AA;

Query Match 55.6%; Score 40; DB 22; Length 261;
Best Local Similarity 53.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGEQE 13
|||: :|||| :
Db 137 lpaiatvspgead 149

RESULT 11
ABG24698
ID ABG24698 standard; Protein; 306 AA.
XX
AC ABG24698;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #24689.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
DR N-PSDB; AAS88885.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 55057; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 306 AA;

Query Match 55.6%; Score 40; DB 22; Length 306;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
|| ::|||||
Db 116 lpgllglspg 125

RESULT 12
AAB92530
ID AAB92530 standard; Protein; 358 AA.
XX
AC AAB92530;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:10688.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10688; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 358 AA;

Query Match 55.6%; Score 40; DB 22; Length 358;
Best Local Similarity 58.3%; Pred. No. 1.8e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQE 13
||: | ||:||
Db 213 pamagpgpgdqe 224

RESULT 13
AAB57055
ID AAB57055 standard; Protein; 668 AA.
XX
AC AAB57055;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1633.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF16258.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 2091-2094; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 668 AA;

Query Match 55.6%; Score 40; DB 21; Length 668;
Best Local Similarity 72.7%; Pred. No. 3.6e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VVGLSPGEQEY 14
|||||||: |
Db 27 vvglspgqyfy 37

RESULT 14
AAY41875
ID AAY41875 standard; Peptide; 18 AA.
XX
AC AAY41875;
XX
DT 09-DEC-1999 (first entry)
XX
DE Rheumatoid arthritis diagnostic protein isoform peptide #26.
XX
KW Human; rheumatoid arthritis; RA; diagnosis; RPI; RADF; detection;
KW rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
KW rheumatoid arthritis diagnostic protein isoform; screening;
KW expression reference protein isoform; prognosis.
XX
OS Homo sapiens.
XX
PN WO9947925-A2.
XX
PD 23-SEP-1999.
XX
PF 15-MAR-1999; 99WO-GB00763.
XX
PR 13-MAR-1998; 98GB-0005477.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Parekh RB, Patel TP, Townsend RR;
XX
DR WPI; 1999-571871/48.
XX
PT Diagnosis of human rheumatoid arthritis by two-dimensional
PT electrophoresis -
XX
PS Disclosure; Page 18; 157pp; English.
XX
CC A method has been developed for the diagnosis of human rheumatoid
CC arthritis (RA) using two-dimensional electrophoresis to generate a
CC two-dimensional array of features. The method can be used for screening,
CC diagnosis and prognosis of RA in a subject or for monitoring the effect
CC of an anti-RA drug or therapy administered to a subject. The method
CC comprises: (a) analysing a sample of serum or plasma and optionally
CC synovial fluid by two-dimensional electrophoresis, to generate a two-
CC dimensional array of features; (b) identifying at least one chosen
CC feature whose relative abundance correlates with the presence or absence
CC of RA; and (c) comparing the abundance of each chosen feature in the
CC sample with the abundance of that chosen feature in serum or plasma from
CC one or more persons without RA, where the relative abundance of the
CC chosen feature or features in the sample indicates the presence or
CC absence of RA in the subject. The method can also be used in clinical
CC studies for testing drugs for therapy of RA, for purification of RA-
CC diagnostic protein isoforms (RPIs), and for production of antibodies to
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
CC compounds that promote or inhibit their activity, which are then used as
CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
CC protocols. AAY41844 to AAY42100 represent RPI peptides, AAY42101 to
CC AAY42103 represent expression reference protein isoform peptides and
CC AAZ25066 to AAZ25068 represent degenerate probes for RPIs, which are all
CC used in the exemplification of the present invention.
XX
SQ Sequence 18 AA;

Query Match 54.2%; Score 39; DB 20; Length 18;
Best Local Similarity 63.6%; Pred. No. 9.7;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
|| : |||||:
Db 8 patlslspger 18

RESULT 15
AAU25388
ID AAU25388 standard; Peptide; 18 AA.
XX
AC AAU25388;
XX
DT 18-DEC-2001 (first entry)
XX
DE Schizophrenia-Associated Protein Isoform (SPI) peptide #617.
XX
KW Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;
KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.
XX
OS Homo sapiens.
XX
PN WO200162785-A2.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-GB00792.
XX
PR 24-FEB-2000; 2000GB-0004415.
PR 28-NOV-2000; 2000US-0750395.
XX
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
PI Herath HMAC, Parekh RB, Rohlff C, Terrett JA, Tyson KL;
XX
DR WPI; 2001-570624/64.
XX
PT New schizophrenia associated protein isoforms and encoding nucleic acid
PT molecules, useful for treatment, diagnosis and prognosis of
PT schizophrenia and screening for potential drugs for treatment and new

PT drug targets -
XX
PS Disclosure; Page 41; 148pp; English.
XX
CC The sequence represents a schizophrenia-associated protein isoform (SPI).
CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable
CC in cerebrospinal fluid, serum or plasma and are useful markers of
CC schizophrenia. The sequences can be used for treatment and diagnosis of
CC schizophrenia, screening, prognosis, monitoring the results of therapy,
CC identifying patients most likely to respond to a particular therapy and
CC identification of new targets for drug treatment. SPI DNA is useful as a
CC nucleic acid probe to detect the presence of nucleic acids or SPIs.
XX
SQ Sequence 18 AA;

Query Match 54.2%; Score 39; DB 22; Length 18;
Best Local Similarity 63.6%; Pred. No. 9.7;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 PAVVGLSPGEQ 12
           || : |||||:
Db       8 patlslspger 18
```

Search completed: September 9, 2002, 06:50:21
Job time: 849 sec